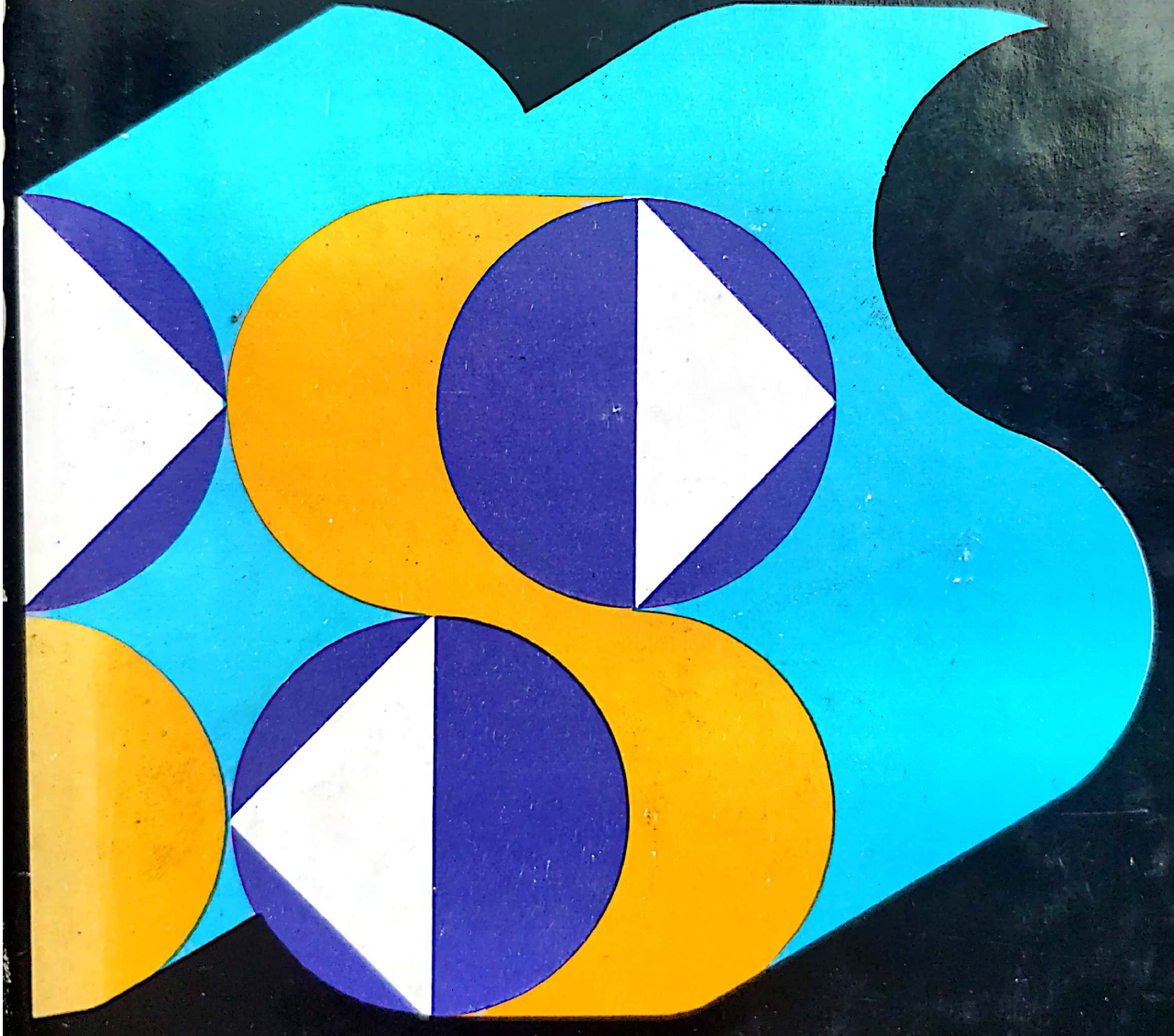
Costache Moineagu | Ion Negură | Veniamin Urseanu

statistica



Editura științifică și enciclopedică

Statistica este destinată celor mai largi și diferite categorii de cititori care, într-un domeniu sau altul, au de-a face cu statistica și cu aplicațiile metodelor sale. Expunerea materiei, riguroasă din punct de vedere științific, este accesibilă grație unei preocupări didactice permanente: demersul este cel firesc, de la simplu la complex, de la concret la abstract, iar prezentarea metodelor este însoțită de exemple practice și de ilustrări menite să consolideze noțiunile practice.

În lucrare este prezentă grija de a valorifica tot ceea ce știința statistică a creat în ultimii ani: de aceea, alături de metodele „clasice", cititorul va găsi elementele cele mai recente care au îmbogățit această știință și i-au diversificat domeniile de aplicație.

Finalitatea *Statisticii* este prin excelență practică și deci este îndreptățită speranța că cititorii vor beneficia din plin, în activitatea lor, de pe urma lecturii cărții.

Scanned with OKEN Scanner

3102

C. MOINEAGU • I. NEGURĂ • V. URSEANU

statistica

Concepte, principii, metode

CUVÎNT ÎNAINTE DE VL. TREBICI

EDITURA ȘTIINȚIFICĂ ȘI ENCICLOPEDICĂ
București, 1976

Scanned with OKEN Scanner

Cuvînt înainte

Cititorul căruia îi sînt familiare manualele, cursurile sau tratatele de statistică, purtînd semnătura unor cunoscuți oameni de știință și care au îmbogățit considerabil literatura ultimelor trei decenii din țara noastră, este în drept să-și pună de la bun început întrebarea, dacă lucrarea de față este un curs sau un manual.

Statistica nu este nici un curs și nici un manual, deși materia este expusă — cititorul va putea aprecia singur acest lucru — după anumite considerente didactice. În intenția autorilor ei — cadre didactice și de cercetare cu îndelungată experiență —, cartea încearcă să prezinte cunoștințele principale de statistică, principiile și metodele sale de bază, într-o concepție coerentă și într-o formă accesibilă, fără vreo restricție impusă de programele analitice. Folosind o formulă elastică, fără nici o concesie în ce privește nivelul științific, *Statistica* se adresează unui cerc larg de cititori interesați de problemele statisticii, căutînd să le ofere o informație cît mai cuprinzătoare, selectată, pe cît este posibil, cu grijă, la nivelul actual al progreselor științei statistice.

Sub o denumire foarte generală, aceea de „domeniul statistic", primul capitol înfățișează problematica generală a statisticii, stăruindu-se îndelung asupra evoluției conceptului de statistică. Excursul istoric — util pentru geneza statisticii și semnificația ei actuală — pune în evidență mutațiile pe care le-a cunoscut statistica, în contextul general al istoriei științelor. Instructivă apare concluzia ce se desprinde din această reiterare a statisticii: în etapa actuală, în care caracterul multidisciplinar s-a accentuat considerabil, statistica rămîne o modalitate indispensabilă pentru studiul fenomenelor aleatoare, indiferent de domeniul din care ele fac parte. Noi și noi discipline au apărut, în ultimele decenii, pe fondul revoluției științifice și tehnice actuale. S-ar fi părut că în aceste condiții statistica și-a pierdut mult din actualitatea sa; mai mult, că ea ar fi devenit anacronică. Lucru surprinzător, statistica a reușit nu numai să se readapteze condițiilor contemporane, dar și să-și renoveze aparatul său metodologic. Cibernetica, teoria sistemelor, modelarea și programarea matematică, în diferitele sale variante, nu numai că nu au luat locul statisticii, ci dimpotrivă, i-au reafirmat importanța, evident într-o optică modernă. Referirile la statistică și modelare, la raportul dintre statistică și informatică, deși succinte, sînt revelatorii în această privință.

Dacă în capitolul întîi, cititorul face cunoștință cu sistemul informațional în accepție modernă, în capitolul al doilea, „Procesul cunoașterii statistice",

Scanned with OKEN Scanner

el va găsi o concluzie cuprinzătoare a ceea ce reprezintă tratarea automată a informației statistice. Considerațiile cu privire la analiza statistică, ca analiză de sistem, sînt de natură să convingă cît de fertilă este abordarea sistemică în statistică ca și, firește, în alte discipline.

Capitolul al treilea, „Probabilități și variabile aleatoare. Repartiții teoretice", al cincilea și al șaselea consacrate „Sondajului statistic", împreună cu metodele de estimație și de verificare a ipotezelor statistice, reprezintă partea de statistică matematică, expusă concentrat. Sînt astfel examinate, pe rînd, noțiunile de probabilități și variabile aleatoare, valorile tipice ale acestora, legile de repartiție și legea numerelor mari. De remarcat că întreaga problematică a teoriei probabilităților, în accepție axiomatică modernă, este însoțită de considerații pertinente de ordin filozofic, pentru a pune în evidență sensul dialectic al probabilității și legii numerelor mari și a infirma unele interpretări neștiințifice. Sondajul statistic beneficiază, de asemenea, de o expunere modernă, care oferă cititorului informațiile cele mai recente din acest domeniu al statisticii matematice.

Capitolul al patrulea, „Repartiții empirice", înfățișează problema centrală a statisticii descriptive: construirea repartițiilor empirice, caracteristicile cantitative și calitative, reprezentarea grafică a repartițiilor, valorile tendinței centrale: medie, mod, mediană, cuartile, dispersie, asimetrie, concentrarea. Economia lucrării justifică tratarea repartițiilor empirice după cele teoretice pe linia demersului „model" — „realitate", cu certe avantaje pentru înțelegerea problemelor.

Capitolul al șaptelea, „Ritm și tendințe", expune problematica seriilor cronologice și a indicilor statistici, într-un mod detaliat, dată fiind importanța acestora în studiile și analiza fenomenelor economice. Nu sînt omise problemele legate de caracterizarea sezonalității fenomenelor, începînd cu metodele cele mai simple și mergînd pînă la cele oferite de analiza spectrală.

Ultimul capitol, al optulea, consacrat regresiei și corelației, cuprinde atît elementele specifice de statistică matematică, cît și cele obișnuite aplicative. Cititorul va regăsi considerații ample în legătură cu premisele calculelor de corelație, care avertizează contra pericolelor ce-l amenință pe cercetător, cînd ele nu sînt strict controlate. Testele de semnificație permit să se asigure validitatea ipotezelor de lucru și a rezultatelor provizorii, dar protecția oferită de ele nu trebuie să dispenseze de reflexia asupra sensului și legitimității concluziilor desprinse din calculele de corelație. O atenție aparte se acordă problemelor particulare puse de corelațiile temporale și extrapolarea funcțiilor de modelare a corelațiilor.

Statistica are un caracter pronunțat metodologic și aplicativ. Considerațiilor speculative și uneori scolastice le-a fost preferată expunerea concretă, însoțită de aplicații ale metodelor statistice, demers socotit mai util pentru cei ce folosesc statistica și care sînt mai puțin interesați în demonstrații și interpretări teoretice. Acesta și este de fapt scopul lucrării: de a pune la dispoziția cititorilor, venind din cele mai variate discipline și domenii, o expunere clară, coerentă a metodelor statisticii și care să înlesnească aplicarea lor în orice domeniu de fenomene, fie ele din societate, fie din natură. Economistul, sociologul, psihologul, geograful, ca și colegii lor din științele naturii, precum și cei interesați pur și simplu de statistică, de unele metode ale acesteia vor putea găsi — sperăm cu multă convingere — informațiile necesare.

Scanned with OKEN Scanner

Statistica nu se substituie manualelor și cursurilor existente. Cu trăsăturile sale specifice determinate de scop și de cuprindere, ea se integrează în evoluția generală a statisticii românești, care datorează mult înaintașilor și autorii ei consideră de datoria lor să-i omagieze.

Statistica românească, inclusiv literatura consacrată problemelor ei, are o îndelungată tradiție, marcată de importante realizări. Înregistrările statistice de tipul catastihelor, catagrafiilor, conscripțiilor, urbariilor, s-au practicat secole de-a rîndul în țara noastră. „Era entuziasmului statistic" — după expresia lui H. Westergaard — este caracteristică și țării noastre: primul recensămînt modern a avut loc în anul 1838, primul oficiu național de statistică s-a înființat în anul Unirii Principatelor, în 1859. Numeroase recensăminte demografice și economice s-au efectuat în decurs de mai bine de un secol. Învățămîntul statistic are de asemenea rădăcini trainice în țara noastră. Literatura statistică a cunoscut încă la mijlocul secolului al XIX-lea cîteva realizări memorabile: ne gîndim în special la „Povățuirile..." lui Ion Ionescu de la Brad.

O etapă calitativă nouă este marcată de școala românească de teoria probabilităților și de statistică matematică, întemeiată în deceniul al treilea și legată de numele academicienilor Octav Onicescu și Gheorghe Mihoc, cu contribuții larg recunoscute pe plan internațional. De certă valoare teoretică și metodologică este cartea prof. N. St. Georgescu-Roegen, intitulată „Metoda statistică" (1933), de o surprinzătoare modernitate.

În noile condiții inaugurate de August 1944, statistica românească a cunoscut o puternică dezvoltare atît pe plan instituțional, cît și pe tărîm științific și didactic. Este un prilej de adîncă satisfacție de a consemna o activitate multilaterală desfășurată de Direcția Centrală de Statistică, de instituțiile noastre de învățămînt superior — matematic, economic, juridic, filosofic, tehnic, medical—în difuzarea cunoștințelor de statistică și în pregătirea temeinică a specialiștilor — în special la Facultatea de cibernetică economică și statistică de la Academia de Studii Economice și la Facultatea de matematică de la Universitatea din București.

Statistica, alăturîndu-se literaturii actuale, va contribui la satisfacerea interesului crescînd manifestat de publicul nostru pentru problemele de statistică, răsplătind în acest fel eforturile depuse de autori.

VL. TREBICI

Scanned with OKEN Scanner

Tabla de materii

Cuvînt înainte de Vl. Trebici *5*

CAPITOLUL 1

Domeniul statisticii *13*

1.1 Conceptul de statistică *13*
1.1.1 De la statistica descriptivă la aritmetica „politică" *14*
1.1.2 Aportul teoriei probabilităților. Quételet și începuturile statisticii moderne *19*
1.1.3 Alte concepții post-quételetiene *22*
1.1.4 Gîndirea statistică românească în secolul al XIX-lea *23*
1.1.5 Statistica modernă *25*
1.1.6 Cîteva concluzii *27*
1.2 Fenomenele de masă ca obiect de investigație statistică *29*
1.3 Demersul statistic *31*
1.4 Statistica în condițiile revoluției științifice și tehnice contemporane *33*
1.4.1 Statistica și noile discipline științifice *33*
1.4.2 Progrese recente în statistică *36*
1.4.3 Statistica și modelarea *37*
1.4.4 Statistica și informatica *37*

CAPITOLUL 2

Procesul cunoașterii statistice *40*

2.1 Informația și sistemul informațional statistic *40*
2.2 Observarea statistică *43*
2.3 Prelucrarea materialului statistic *44*
2.4 Analiza statistică și dezvoltarea ei ca analiză de sistem *58*

CAPITOLUL 3

Probabilitate și variabilă aleatoare. Repartiții teoretice *61*

3.1 Teoria probabilităților. Concepte *62*
3.2 Variabile aleatoare unidimensionale și repartițiile lor *72*
3.2.1 Noțiuni introductive. Definiții *72*

Scanned with OKEN Scanner

3.2.2 Repartiția variabilei aleatoare discrete. Poligonul de repartiție *74*
3.2.3 Repartiția variabilei aleatoare continue. Funcție de repartiție *76*
3.2.4 Densitate de repartiție *79*
3.3 Valori tipice ale variabilelor aleatoare *82*
3.3.1 Valori tipice ale variabilelor aleatoare discrete *82*
3.3.2 Valori tipice ale variabilelor aleatoare continue. Valori tipice de poziție *90*
3.4 Repartiții ale unor variabile aleatoare *94*
3.4.1 Repartiții ale unor variabile aleatoare de tip discret *94*
3.4.2 Repartiții ale unor variabile aleatoare de tip continuu *99*
3.5 Repartiții multidimensionale *109*
3.5.1 Generalități. Conceptul de sistem de variabile aleatoare *109*
3.5.2 Repartiția sistemului de n variabile aleatoare *110*
3.5.3 Valori tipice ale unui sistem de variabile aleatoare *111*
3.5.4 Repartiție multidimensională normală. Repartiție condiționată *113*
3.6 Legea numerelor mari. Teorema limită centrală *116*
3.6.1 Legea numerelor mari *116*
3.6.2 Teorema limită centrală *119*

CAPITOLUL 4

Repartiții empirice *120*

4.1 Construirea repartițiilor empirice. Reprezentare grafică *120*
4.1.1 Cazul unei caracteristici calitative *121*
4.1.2 Cazul unei caracteristici cantitative *122*
4.2 Descrierea numerică a repartițiilor empirice *128*
4.2.1 Parametrii de nivel *128*
4.2.2 Parametrii dispersiei *141*
4.2.3 Caracteristicile formei *149*
4.2.4 Curba de concentrare *150*

CAPITOLUL 5

Sondajul statistic (I). Metode de estimație și de verificare a ipotezelor statistice. Eșantioane mari *152*

5.1 Eșantioane. Populație originară *152*
5.2 Avantajele și obiectivele sondajului *153*
5.3 Tipuri de sondaj *154*
5.3.1 Sondaj pur aleator *154*
5.3.2 Sondaj dirigat *156*
5.3.3 Sondaj mixt *156*
5.4 Sondaj simplu din populații cu o singură caracteristică cantitativă *156*
5.4.1 Generalități. Repartiții de sondaj *156*
5.4.2 Estimații în teoria eșantioanelor de mare volum *158*
5.4.3 Eroarea standard a repartiției de sondaj a mediei aritmetice *162*
5.4.4 Cîteva observații cu privire la repartiția de sondaj a mediei aritmetice *163*
5.4.5 Construirea intervalelor de încredere în cazul repartițiilor normale *164*
5.5 Criterii și teste de apreciere a semnificației statistice a valorilor tipice de sondaj *165*

Scanned with OKEN Scanner

5.5.1 Ipoteza diferenței nule *165*
5.5.2 Criterii de semnificație bazate pe repartiții normale *167*
5.6 Eroarea standard a medianei *171*
5.7 Erorile standard ale momentelor de ordin superior și ale altor valori tipice de sondaj *173*
5.7.1 Erorile standard ale momentelor convenționale și centrate *173*
5.7.2 Eroarea standard a diferenței dintre două abateri medii pătratice de sondaj *175*
5.7.3 Eroarea standard a coeficientului de variație *176*
5.8 Sondaj tipic (stratificat) *177*
5.8.1 Sondaj tipic simplu *177*
5.8.2 Sondaj tipic proporțional *181*
5.8.3 Sondaj tipic optim *183*
5.9 Sondaj în faze *185*
5.9.1 Sondaj în două faze *186*
5.9.2 Eroarea standard a mediei în cazul sondajului în două faze *190*
5.10 Sondaj simplu privind o caracteristică-atribut *191*
5.10.1 Considerații introductive *191*
5.10.2 Media și eroarea standard a caracteristicii-atribut în cazul sondajului simplu *192*
5.10.3 Compararea a două proporții de sondaj ale unei caracteristici-atribut oferite de două eșantioane diferite *195*
5.10.4 Abateri de la condițiile sondajului simplu și efectul lor asupra erorii standard a proporției cazurilor favorabile *198*
5.11 Tehnici particulare de sondaj *200*
5.11.1 Sondaj de serii *200*
5.11.2 Sondaj mecanic (sistematic) *202*
5.11.3 Sondaj secvențial *203*
5.12 Dimensionarea eșantioanelor *205*

CAPITOLUL 6

Sondajul statistic (II). Sondaje pe bază de eșantioane mici. Concluzii *208*

6.1 Considerații introductive *208*
6.2 Estimarea valorilor tipice ale populației originare pe bază de eșantioane mici *208*
6.3 Cîteva observații privitoare la estimare și gradele ei de libertate *210*
6.4 Criterii și teste de apreciere a semnificației statistice a estimațiilor pe bază de eșantioane mici *211*
6.4.1 Valoarea gnoseologică a testelor de semnificație *211*
6.4.2 Ipoteza de normalitate. Aplicații ale repartiției Student *212*
6.5 Compararea valorilor medii de sondaj în cazul a două eșantioane mici *216*
6.5.1 Ipoteza de egalitate a dispersiilor necunoscute *216*
6.5.2 Ipoteza de inegalitate a dispersiilor necunoscute *218*
6.6 Cîteva observații generale pe marginea aplicațiilor referitoare la eșantioane mici *222*
6.7 Cîteva precizări și concluzii finale privind sondajul statistic *223*
6.7.1 Problema estimării *223*
6.7.2 Conceperea planului de sondaj și problema omogenității populației originare *226*

Scanned with OKEN Scanner

CAPITOLUL 7

Ritm și tendințe *228*

7.1 Seriile cronologice în metodologia statistică *228*
7.1.1 Caracteristici generale *228*
7.1.2 Clasificarea și reprezentarea grafică a seriilor cronologice *231*
7.1.3 Indicatori și metode de reflectare a fenomenelor sociale *236*
7.2 Indicatori absoluți și relativi. Calculul ritmului mediu de dezvoltare *238*
7.2.1 Indicatori absoluți și relativi *238*
7.2.2 Metode de determinare a ritmului mediu *240*
7.3 Metoda indicilor statistici *257*
7.3.1 Caracteristici, retrospectivă, clasificări *257*
7.3.2 Tehnica de calcul al indicilor *262*
7.3.3 Indicii de structură și descompunerea influențelor pe factori *276*
7.3.4 Indicii teritoriali *282*
7.4 Calculul statistic al tendinței de dezvoltare și măsurarea oscilațiilor sezoniere *287*
7.4.1 Procedee de ajustare a seriilor cronologice *287*
7.4.2 Calculul variațiilor sezoniere *301*

CAPITOLUL 8

Corelație *308*

8.1 Considerații preliminare *308*
8.2 Conceptul de corelație *311*
8.3 Premisele calculelor de corelație *313*
8.4 Regresia *320*
8.4.1 Modelul regresiei simple *322*
8.4.2 Modelul regresiei multiple *332*
8.5 Coeficientul de corelație. Raportul de corelație *339*
8.5.1 Coeficientul de corelație *339*
8.5.2 Raportul de corelație *344*
8.6 Corelații temporale *346*
8.7 Măsurarea influenței factorilor în relația stohastică *350*
8.8 Previziunea pe baza funcției de modelare a corelației *352*
Anexe (1—10) *355*
Bibliografie *373*
Indice de nume *379*
Indice de materii *383*

Scanned with OKEN Scanner

CAPITOLUL 1
Domeniul statisticii

1.1 CONCEPTUL DE STATISTICĂ

Cu greu s-ar putea găsi alt concept care să fi suferit, în evoluția lui istorică, atîtea schimbări în conținut și care să fie atît de mult controversat ca statistica. Drumul parcurs de statistică este același cu al tuturor științelor: de la practică la teorie.

Cuvîntul *statistică*, precum și primele formulări ale unui concept de statistică, datează din sec. al 18-lea, dar modul statistic de abordare a fenomenelor își are originea în cele mai vechi timpuri. Istoria consemnează operații de numărare a populației, de măsurare a terenurilor sau a altor resurse la popoarele antice [1]. Asemenea operații, indispensabile pentru ridicarea armatelor, așezarea impozitelor și evaluarea bogățiilor, cunoscute bine și în *polis*-urile Greciei antice, au culminat cu recensămintele *(census)* și anchetele romane [2]. În imperiul roman s-a organizat și o evidență a mișcării naturale a populației [3].

[1] Egiptenii cunoșteau numărători ale populației, cu caracter regulat: la intervale de doi ani în vremea regatului vechi (2778—2160 î.e.n.) și anuale cu începere de la a 6-a dinastie. Este atestat un recensămînt pe vremea faraonului Amenophis III (1408—1372), care surprinde prin ingeniozitate. Chinezii dispuneau de informații despre populație și pămînturi încă de pe timpul legendarului împărat Yu (pe la anul 2300 î.e.n.), inserate în *Shu-King* (*Cartea istoriei* — culegere de cele mai vechi documente scrise ale Chinei). Pentru statele antice (Lagash, Ur, Nippur, Akkad) din ținutul aflat între Tigru și Eufrat, tăbliță de lut („cărțile de lut") au conservat date cu privire la agricultură, creșterea animalelor, construcții de corăbii, navigație etc. (mileniul al 3-lea).

Și alte state trebuie să fi efectuat numărători ale populației și cadastre, mai cu seamă dacă dispuneau de armate mari, ridicau construcții publice sau repartizau terenurile [36; 46; 109].

[2] Tradiția antică atribuie instituirea *census*-ului regelui Servius Tullius (sec. 6 î.e.n.). Cu ocazia *census*-ului, fiecare cetățean trebuia să declare numele și prenumele, vîrsta, familia sa, numărul sclavilor și bunurile de orice natură. După cum ne informează Titus Livius, *census*-ul a permis stabilirea obligațiilor civile și militare ale cetățenilor nu numai pe cap de om, ca mai înainte, ci și după avere *(tributum ex censu)*. O practică mai evoluată și mai complexă a unei astfel de investigații a avut loc pe vremea împăratului August (anii 27 î.e.n. și 14 e.n.). P.C. Tacitus pomenește de o vastă anchetă, care a cuprins cetățenii și aliații aflați sub arme, navele, resursele de toate felurile, veniturile publice. Bilanțul bogățiilor imperiului, scris cu mîna lui August, a fost prezentat senatului de succesorul său. Potrivit relatărilor lui L.C. Lactantius, împăratul Traian, imediat după cucerirea Daciei, a ordonat efectuarea unui recensămînt general al noii provincii romane *(census provinciae)*, după formula bine cunoscută în imperiu [36; 93].

[3] Sub Antonin Piul (138—161), declararea nașterilor a devenit obligatorie și se făcea în Roma la prefectul tezaurului, iar în provincie, la *tabularium* (birouri speciale în care se păstrau registrele fiscale, vamale, cadastrele etc.). Administrația romană a organizat în Dacia atari *tabularium*, despre care avem mărturii de la Apulum și Sarmisegetuza [36; 127].

Scanned with OKEN Scanner

Investigații analoage au continuat în Evul Mediu. Caracteristice sînt inventarierile de bunuri ale feudelor, mînăstirilor, corporațiilor sau breslelor și însemnările curente ale nașterilor, căsătoriilor, decèselor la parohii, premergătoare actelor de stare civilă. N-au lipsit lucrări de evidență din inițiativa statului [4].

Contabilizarea resurselor umane și materiale în lumea antică și cea medievală a izvorît dintr-o gîndire practică, utilitară și s-a executat, desigur, fără criterii riguroase, în mod empiric. Informațiile se culegeau mai cu seamă în scopuri fiscale, militare ori administrative, cu mijloace variate, mergînd de la nodurile făcute pe sfori de culori diferite din timpul incașilor pînă la cadastrul englez din epoca lui William Cuceritorul. Tipurile de colectivități observate și modalitatea de descriere fac parte, incontestabil, din domeniul atribuit statisticii; nu se găsește însă în acele vremuri „vreun principiu de tehnică, nici vreo idee precisă despre natura științifică a statisticii" [61, p. 13].

Din antichitate și Evul Mediu s-au păstrat înregistrări care și astăzi surprind prin ingeniozitatea lor și s-ar putea ca noi cercetări istorice să arate că statistica de atunci reprezenta mai mult decît simple colecții de informații fragmentare. Lucrările despre incași, de exemplu, ne dezvăluie „descrieri foarte bine articulate și calcule în vederea unui efort nebănuit de planificare" [5]. Oricum, chiar la acest nivel de dezvoltare, „statistica veche" a servit numai nevoilor practice de orientare. Depășirea stadiului de statistică practică sau materială în exclusivitate s-a produs în momentul în care, peste interesele practice imediate, s-au evidențiat interese de ordin științific, de cunoaștere critică a fenomenelor înregistrate. Înregistrările empirice au folosit disciplinei statistice, la începutul ei, ca instrument de vehiculare a unui nou domeniu de cunoaștere și ca sursă documentară pozitivă, oferindu-i material de analiză și interpretare.

1.1.1 De la statistica descriptivă la aritmetica „politică"

Într-o primă concepție, statistica echivala cu *descrierea statului*, expunerea situației geografice, economice, politice. Acest gen de statistică, cu modelul său mai îndepărtat în *Politica* lui Aristotel [6], a fost cultivat mai întîi de italieni. Încă din sec. al 13-lea și al 14-lea, celebrele *relazioni* (rapoarte) ale republicii venețiene concentrau, grație negustorilor și apoi agenților diplomatici, multiple informații despre condițiile partenerilor comerciali. Se crede că pe

[4] Cităm, de exemplu, lucrările: *Breviarium fiscalium* din epoca lui Carol cel Mare (771—814), în care sînt minuțios descrise locuințele, uneltele de casă, proviziile, animalele de curte, speciile de culturi și *Doomesday Book* sau *Liber iudiciarius seu censualis Willelmi I regis Angliae*, care, în afară de enumerarea locuitorilor, include descrierea statistică a patrimoniilor, veniturilor, pămînturilor, servituților prediale (funciare). Cartea lui William Cuceritorul, compusă prin anii 1083—1086, este considerată prima însemnare censuară de mare proporție după *census*-ul roman [46].

[5] Lucrările lui Louis Baudin la care se referă André Piatier în [109, p. 47]. Cu mult înainte, istoricii cuceririi țării incașilor de către conchistadorii spanioli și-au exprimat uimirea față de evidențele destul de complexe descoperite la un popor care nu avea ca mijloace de scriere și calcul decît sfori *(khipu)* de culori variate, înnodate și combinate în chip diferit [93].

[6] Aristotel (384—322 î.e.n.), „minte enciclopedică a antichității" (Marx), a descris, în ale sale πολιτεῖαι, 158 constituții politice, din care a rămas doar un singur fragment relativ la constituția Atenei.

Scanned with OKEN Scanner

această cale „s-au acumulat materiale comparabile cu acele care se găsesc în zilele noastre în anuarele de comerț exterior" și au fost utilizate în politica comercială oficială, cu tot „conținutul lor incert și non-numeric" [55].

Lucrările de descriere a statelor se înmulțesc în epoca Renașterii. Printre cele mai reprezentative amintim *Del governo e amministrazione di diversi regni e repubbliche* (F. Sansovino) și *Le relazioni universali* (G. Botero) [7]. Lui F. Sansovino i se atribuie meritul de a fi luat în considerare, în cartea sa, descrierea numerică a statelor: bogățiile, populația, comerțul, industria și finanțele lor. Cei doi autori au avut numeroși continuatori în alte țări și îndeosebi în Germania.

Dezvoltarea etatismului mercantilist european antrenează o intensificare a preocupărilor pentru cunoașterea situației a cît mai multor țări. În întîmpinarea cererii sporite de informații a apărut faimoasa colecție *Respublicae Elzevirianae*, un fel de enciclopedie a statelor [8].

Curentul „descrierea statului" a atins apogeul în sec. al 17-lea și al 18-lea, cînd, în Germania, s-a constituit o adevărată școală, cunoscută sub denumirea de *școala descriptivă germană.*

Fondatorul școlii descriptive germane a fost Hermann Conring (1606—1681), profesor la Universitatea din Helmstedt. Prin originea sa olandez, Conring a cunoscut pe Jan de Laet și a adus o contribuție prețioasă la realizarea bibliotecii elzeviriene, care a format dealtfel și baza cursului său universitar despre știința descriptivă a statului. Prelegerile lui Conring au fost, fără voia sa, publicate de un discipol, în anul 1671, sub titlul *Notitia rerum publicarum* sau *Descriptio singularis rerum publicarum aevi nostri*. Ca cel dintîi manual de știință a statului se citează însă lucrarea *Teutscher Fürstenstaat* (1656), elaborată de Veit Ludwig von Seckendorff (1626—1692), profesor la Universitatea din Halle. Martin Schmeitzel (1679—1747) a introdus, la Universitatea din Halle, cursuri despre *Știința camerală* și *Notitia rerum publicarum*, apoi a ținut un curs intitulat *Collegium politico-statisticum*. Gottfried Achenwall (1719—1772) continuă învățămîntul denumit *Staatskunde* la Universitatea din Göttingen. Prin învățămîntul pe care-l asigurau, profesorii universitari erau în mod automat consilieri în gestiunea afacerilor statului și responsabili de formarea funcționarilor necesari administrației [46; 55].

Descrierea statului, pînă atunci vulgară, a devenit disciplină de predare academică, încadrată într-un sistem construit după norme teoretice și practice, care s-au elaborat și dezvoltat în universitățile germane. Noua disciplină *(Staatskunde)*, derivată din literatura în materie existentă și nu din cercetări originale ale autorilor ei, a primit ulterior numele de statistică *(Statistik)* [9].

[7] Lucrarea lui F. Sansovino (1521—1583) cuprinde descrierile a 22 state, începînd cu Sparta, Atena, Roma etc. și terminînd cu statul imaginar Utopia. În chip analog, G. Botero (1533—1617), considerat unul din precursorii demografiei și căruia îi datorăm relații despre Moldova, se referea la o serie de state; la el apar rudimentele unei statistici comparate. Cam în aceeași vreme (a doua jumătate a sec. al 16-lea), L. Guicciardini publică *Descrittione di tutti i paesi bassi altrimenti detti Germania inferiore*, la Anvers, transplantînd noua disciplină în Țările de Jos. În Franța, E. Pasquier a scris *Recherches de la France*, iar mai tîrziu Pierre d'Avity a întocmit lucrarea *Les etats, empires et principautés du monde*, asemănătoare lucrării numită *Kosmographie* a lui S. Münster, din prima jumătate a sec. 16-lea [55].

[8] După numele ilustrei familii de tipografi Elzevir din Olanda. Colecția a însumat cîteva zeci de volume — cele mai multe publicate între anii 1625 și 1652 — și a avut ca principal editor pe Jan de Laet.

[9] De obicei, se atribuie lui G. Achenwall prima întrebuințare a termenului de statistică în lucrarea *Abriss der Staatswissenschaft der europäischen Reiche* (1749). Adjectivul *statisticus* apăruse mai devreme în lucrări scrise în limba latină, ca de exemplu: *Microscopium statisticum quo status imperii Romano-Germanici repraesentatur* (1672) a lui Helenus Politanus; *Itinerarium Germanice politicum* (1675) de Phil. Andr. Oldenburg (în care se vorbește de *rationes statisticae*). Adesea se acordă lui Achenwall nu numai paternitatea denumirii, ci și a disciplinei, onoare pe care el n-a revendicat-o niciodată și care pare exagerată. „A declara pe Achenwall « părintele

Protagoniștii școlii germane s-au degajat, într-o oarecare măsură, de orientarea pur descriptivă a predecesorilor. G. Achenwall a definit statistica *ex professo*: „Ansamblul caracteristicilor statale reale ale unui imperiu sau ale unei republici formează, în sens larg, constituția lor și învățătura despre constituția unuia sau a mai multor state este statistica (Staatskunde) sau descrierea statului". În aceeași concepție, folosul principal al statisticii constă în faptul că „prin ea sîntem puși în situația nu numai de a judeca corect și temeinic felurite treburi de stat, ci și de a dobîndi iscusința de a le trata, atunci cînd este cazul, în mod deliberat și concret" [10]. Statistica avea să înfățișeze deci, după Achenwall, situația *(status)* a tot ceea ce era remarcabil, esențial, într-un stat, la o anumită etapă. Și remarcabile — a precizat apoi A. L. Schlözer — erau toate lucrurile de natură să influențeze fericirea sau nefericirea locuitorilor, prosperitatea sau mizeria lor, forța sau slăbiciunea poporului [11].

Termenul „statistică", definit în spiritul școlii din Göttingen, se întîlnește la mulți autori de renume ai epocii. „Sînt aproape patruzeci de ani — scria E. A. Zimmerman într-un studiu apărut în 1787 — de cînd o ramură a cunoașterii politice care are ca obiect puterea actuală și relativă a statelor moderne... a devenit disciplină distinctă... În forma convenabilă pe care a căpătat-o acum... această disciplină, desemnată prin noul nume de *statistică*, a devenit un obiect de studiu favorit în Germania" [12]. Pe urmă, John Sinclair, adoptînd o definiție ceva mai largă, afirma drept scop al statisticii „de a constata suma fericirii de care se bucură o populație și mijloacele de a o mări" [13]. După cunoscutul economist francez J. B. Say (1767—1832), statistica „expune situația producțiilor și consumațiilor unei sau mai multor națiuni, la o epocă desemnată sau la mai multe epoci succesive, precum și starea populației sale, forțelor sale, actelor ordinare care se petrec și care pot să se supună la aprecierea calculului" [114, p. V].

statisticii » — scrie M. Block — înseamnă a face un al doilea Amerigo Vespucci, care dă numele său unui continent pe care nu l-a descoperit". Însuși numele de statistică se crede popularizat de către Achenwall, însă inventat de profesorul său Schmeitzel, originar din Brașov. Etimologic, cuvîntul statistică provine de la *status*, care înseamnă *stare*, dar și *stat*, pentru că însuși statul este o stare — *status reipublicae* — starea comunității [116].

Statistica descriptivă s-a numit și *statistică universitară*, deoarece reprezentanții săi erau învățați de la universități (mai ales germane).

[10] Citat din ediția a 6-a a lucrării (vezi *supra*), intitulată: *Staatsverfassung der heutigen vornehmsten europäischen Reiche und Völker im Grundrisse*, Göttingen, 1781, p. 4—5, reproduse în [120].

[11] A.L. Schlözer (1735—1809), unul dintre cei mai fideli discipoli ai lui Achenwall și succesorul său la catedra universitară, a exprimat „esența statului" și, ca atare, obiectul statisticii, în trei cuvinte: *Vires unitae agunt*. Prin *vires* a înțeles forțele naturale și dobîndite (oamenii, teritoriul și avantajele naturale), *unitae* — forma și organele de guvernămînt, iar *agunt* — guvernarea și administrația. Lui îi aparține aforismul: „Statistica este istorie în repaus, istoria — statistică în mișcare". Citat după [61].

[12] E.A. Zimmerman, profesor de filozofie a naturii la Braunschweig, a scris un studiu politic al situației Europei în limba engleză: A political Survey of the Present State of Europe (1787) [134, p. 17].

[13] Citat după [41, p. 12]. John Sinclair a fost editorul și organizatorul primului volum al lucrării *Statistical Account of Scotland*. Cuvîntul „statistică" pe care l-a întrebuințat a părut cu totul nou în Anglia: „... mulți oameni — spunea el — au fost întîi surprinși că utilizez cuvinte noi ca *statistică* și *statistic*, presupunînd că un alt termen din limba noastră ar avea același sens... și deoarece am crezut că un cuvînt nou ar atrage mai mult atenția publicului, am hotărît să-l adopt și cred că, acum, acest termen este complet naturalizat și încorporat în limba noastră". A se vedea [134, p. 17].

În decursul jumătății de secol care a urmat nașterii noii discipline, structura studiilor de descriere a statului a început să se modifice în sensul consolidării *determinărilor numerice*. Un reputat geograf, A. F. Büssching, a creat chiar o specie de statistică comparativă, inspirat, se pare, de lucrarea istoricului și filologului danez J. P. Anchersen, intitulată *Descriptio statuum cultiorum in tabulis* (1741). Inovația lui Büssching a fost proclamată drept excelentă de însuși Schlözer „păstrător intratabil al tradițiilor maestrului său Achenwall". El a recunoscut, în cele din urmă, că o statistică fără cifre nu-i nimic: „Indicații generale ca acestea: țara produce un vin *delicios*, ea are *bogate* manufacturi, un comerț *înfloritor*, *puțin* grîu, ne sînt furnizate de toate geografiile; dar aceste indicații, atît timp cît nu sînt precizate prin cifre, nu servesc la nimic" *.

Pătrunderea mai puternică a elementului numeric, favorizată de evidențe oficiale mai abundente și mai ușor accesibile, nu a schimbat conținutul statisticii, așa cum a fost conceput de autorii germani din sec. al 18-lea. Statistica a intrat în istoria disciplinelor științifice ca o disciplină descriptivă, orientată spre *informație*, și nu spre căutarea de legi, cu un mod de expunere preponderent verbal.

În cadrul gîndirii și practicii școlii descriptive germane au fost concepute și elaborate lucrări de descriere a statelor în diferite țări europene. În Țările Române, prima și cea mai reprezentativă lucrare a genului este *Descrierea Moldovei* a lui Dimitrie Cantemir. Lucrarea — o expunere monografică pe plan geografic, politic, economic, social și cultural — s-a impus în mod deosebit atenției contemporanilor, așezîndu-l pe autor printre fruntașii statisticii descriptive europene [14].

Pe vremea constituirii statisticii ca disciplină descriptivă a statului, se năștea în Anglia, în afara universităților, o statistică cu totul deosebită, sub numele de *aritmetica politică*. Denumirea însăși este relevatoare pentru semnificația atașată conceptului de statistică: analiza datelor de observație prin procedee matematice, desprinderea regularităților în fenomenele social-economice și chiar formularea de previziuni. Folosirea metodei analitice, preconizată de filozofii materialiști metafizicieni ai timpului, recurgerea la instrumentul matematic și căutarea legităților marcau un substanțial progres, prefigurînd statistica modernă.

Aritmeticienii politici englezi, în frunte cu John Graunt, William Petty, Edmund Halley, au pus accentul mai ales pe studiul fenomenelor demografice, deoarece vedeau în populație izvorul bogăției și puterii Angliei. Cercetînd cu atenție registrele de nașteri și decese ale Londrei, Graunt [15] s-a străduit să descopere legitățile creșterii populației, ale echilibrului numeric dintre sexe, ale fertilității, mortalității etc. El a atras atenția, pentru întîia dată,

* M. BLOCK, *Traité theorique et pratique de statistique*, Paris, 1878, p. 89.

[14] *Descrierea Moldovei* a fost redactată la cererea Academiei de științe din Berlin, al cărei membru era Dimitrie Cantemir, din 1714. Scrisă în latinește, pe la 1716, cu titlul *Descriptio antiqui et hodierni status Moldaviae*, lucrarea a rămas în manuscris pînă la tipărirea ei în limba germană, în 1769—1770. A fost tradusă apoi în rusește și grecește, iar prima versiune în limba română s-a tipărit în 1825, la Mînăstirea Neamțului [20].

[15] John Graunt (1620—1674), negustor de stofe și comandant al miliției londoneze, cu modestă cultură, dar cu o intuiție genială, a inițiat „explorarea științifică a unui cîmp fecund, pentru cei dinainte absolut necunoscut" [46], din care s-a născut lucrarea *Natural and Political Observations upon the Bills of Mortality* (1662), considerată ca marcînd începutul demografiei, ca știință.

Scanned with OKEN Scanner

asupra predominanței nașterilor masculine în raport cu nașterile feminine (107 băieți la 100 fete), a calculat primele rate de mortalitate pe grupe de vîrstă și a alcătuit — în ipoteza unei populații staționare — o rudimentară tabelă de mortalitate.

În explorările și în popularizarea descoperirilor sale, Graunt a fost sprijinit de amicul său William Petty [16], un „cugetător plin de cutezanță", „părintele economiei politice și într-o oarecare măsură inventatorul statisticii", cum l-a caracterizat K. Marx. Petty, creator al denumirii *political arithmetic*, a invitat, într-o operă cu același titlu, la studiul fenomenelor social-economice „prin intermediul numerelor, al greutăților și al măsurilor..." și a folosit estimațiile cantitative în compararea potențialului uman și economic al diferitelor state și, în particular, în evaluarea posibilităților comerciale și maritime ale Angliei în rivalitatea sa cu Franța.

Cercetări interesante în domeniul aritmeticii politice a întreprins Halley [17]. De numele lui sînt legate estimări ale populației cîtorva state, întemeiate pe criterii mai raționale decît cele adoptate anterior, alcătuirea primei tabele complete de mortalitate, introducerea ideii de durată probabilă a vieții, cu aplicație la calcularea rentei viagere anuale în funcție de vîrstă.

Alți doi reprezentanți de seamă ai școlii engleze, Davenant și King, au contribuit la promovarea teoriei și practicii aritmeticii politice în investigarea fenomenelor economice. Pentru Charles Davenant, aritmetica politică era „arta de a judeca, cu ajutorul cifrelor, lucrurile care privesc guvernul" [18]. Gregory King s-a dedicat calculelor privind veniturile statului, pe calea impozitelor, din agricultură, clădiri, bunuri mobile; de asemenea a încercat să stabilească mărimea capitalurilor de care dispunea Anglia în sec. al 17-lea, pierderile suferite în urma războaielor, incendiilor, epidemiilor. În preocupările lui King a intrat și populația Angliei, acordînd atenție studiului veniturilor, cheltuielilor și acumulărilor realizate de diferite categorii de familii [19].

Școala aritmeticii politice engleze și-a extins influența și în alte țări. Curînd, a apărut o literatură aritmetică politică europeană, avînd repercusiuni profunde asupra curentului statistic universitar. În Germania, aritmetica politică a obținut un succes decisiv, în lupta aprigă cu școala descriptivă din Göttingen, prin aportul lui J. P. Süssmilch [20]. Beneficiar al literaturii colectate de

[16] William Petty (1623—1687), un mercantilist curios și original, fost profesor de anatomie la Oxford, a călătorit în Olanda, unde a luat cunoștință de realizările și succesele fundației Elzevir, sub influența cărora s-a aflat. În cercetările efectuate, W. Petty a accentuat aspectele economice. Lucrarea sa *Political Arithmetic* a însemnat începuturile statisticii economice, ca ramură a științelor sociale [55].

[17] Edmund Halley (1656—1742), astronom englez, s-a ocupat și de probleme demografice. Pe baza datelor despre nașterile și decesele în orașul Wroclaw (pe atunci, Breslau), culese de pastorul Kaspar Neumann (1648—1715), cel care a introdus aritmetica politică în Germania, a construit tabelele de mortalitate publicate în 1693. Aceste tabele au fost folosite de Cornelius de Witt, om de stat olandez, pentru a vinde cu mari profituri rente viagere, salvînd astfel, într-un moment critic, finanțele Olandei [13].

[18] În lucrarea *Discourses on the public revenues* (1698).

[19] Prin aplicarea metodelor matematice în analiza economică, la determinarea venitului național, a elasticității prețurilor, Gregory King (1648—1712) apare ca unul din precursorii econometriei.

[20] Johann Peter Süssmilch (1707—1767), „meteor, strălucitor și solitar", cum l-a numit mai tîrziu statisticianul von Oettingen, este considerat, alături de John Graunt, unul din fondatorii demografiei. Lucrarea sa de bază, o carte de statistică a populației, poartă titlul: *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen* (1741).

Academia din Berlin, Süssmilch a reușit să realizeze o sinteză magistrală a aritmeticii politice. În lucrările sale se găsesc elemente ale unei concepții fecunde în rezultate științifice: căutarea explicației fenomenelor vieții umane și indicarea legilor ce le guvernează, corelarea fenomenelor demografice și economice, examinarea fenomenelor în dinamică, identificarea regularităților în masa de observații, scrupuloasa alegere și evaluare critică a datelor.

Aritmetica politică s-a dezvoltat tot mai mult în demografie, care „se mulțumea în acea epocă să fie statistică" [19, p. 56]. Studiile demografice lăsau să se întrevadă că previziunea fenomenelor colective ar putea avea o însemnătate practică; dealtfel, tabelele de mortalitate, întocmite în sec. al 18-lea, au constituit punctul de plecare al înfloritoarei industrii de asigurări [21].

1.1.2 Aportul teoriei probabilităților. Quételet și începuturile statisticii moderne

Așadar, în sec. al 17-lea și al 18-lea, în concepția despre statistică s-au confruntat două curente: statistica înțeleasă ca descriere empirică și verbală a caracteristicilor statului (statistica descriptivă sau universitară) și statistica orientată spre analiza fenomenelor sociale, în căutarea de legități, apelînd constant la date și calcule numerice (aritmetica politică). În disputa, care opunea cele două curente, a triumfat curentul de tendință modernă reprezentat de aritmetica politică [22]. Statistica descriptivă, agonizantă în pragul sec. al 19-lea, a sfîrșit prin a se combina cu statistica practică, deznodămînt pricinuit și de disocierea, pe măsura constituirii în discipline autonome, a celei mai mari părți din domeniile asupra cărora se întindea odinioară.

În timp ce descrierile statistice devin obiect al instituțiilor publice, domeniul conceptual al aritmeticii politice capătă o nouă dimensiune cu introducerea calculului probabilităților. Generat de studiul jocurilor de noroc, calculul probabilităților se dovedise în scurtă vreme un instrument puternic de cercetare a fenomenelor aleatoare [23]. Primele încercări de a utiliza resursele noii științe în analiza statistică a fenomenelor economice și sociale au fost făcute

[21] Antoine Deparcieux (1703—1768) a construit primele tabele franceze de mortalitate, după listele de asigurări (*Les tontines*) și registrele de decese din mînăstirile pariziene, despuiate pe o lungă perioadă (1607—1745). Deparcieux a publicat, în 1746, lucrarea *L'essai sur les probabilités de la durée de la vie humaine.*

Per Wilhelm Wargentin (1717—1783) a elaborat tabele de mortalitate pentru Suedia, pe baza recensămintelor populației (1757; 1760; 1763) și a buletinelor de decese (1757—1763), publicate în 1766. Sînt primele tabele de mortalitate care se referă la populația unui întreg stat.

[22] Curios lucru, deși optica numerică prevalase față de optica verbală, denumirea aritmetică politică a dispărut, substituindu-i-se denumirea de statistică, rezervată de școala germană „descrierii statului" (*Staatskunde*).

[23] Precursor al teoriei probabilităților, Gerolamo Cardano (1501—1576), matematician și filozof italian, publică *Liber de ludo aleae* (1520) — manual pentru jucătorii de noroc. Meritul fundamentării teoriei probabilităților aparține francezilor Blaise Pascal (1623—1662) și Pierre de Fermat (1601—1665). Acestor doi mari geometri „li se datorează primele elemente ale științei probabilităților, a căror descoperire se numără printre lucrurile remarcabile, făcînd cea mai mare cinste minții umane" [Laplace, 64, p. 245]. După ce teoria probabilităților a depășit faza sa embrionară, în timpul căreia a suscitat mai ales „un interes de curiozitate", a trecut un anumit timp pînă să fie aplicată sistematic în statistică.

Scanned with OKEN Scanner

de Bernoulli [24] și mai cu seamă de Laplace [25]. În lucrarea *Ars conjectandi* (1713), Bernoulli a formulat faimoasa „lege a numerelor mari", a deosebit probabilitatea apriori (matematică, logică sau deductivă) de probabilitatea aposteriori (statistică, empirică sau inductivă) și a schițat posibilitățile aplicării calculului probabilităților în economie (parte neterminată). Laplace a adunat, în *Théorie analytique des probabilités* (1812), toate aplicațiile calculului probabilităților, a evidențiat avantajele în studiul fenomenelor dependente de cauze complexe și a demonstrat utilizarea posibilă a acestui calcul în exploatarea rezultatelor statistice privind fenomene sociale. După o apreciere necontestată, Laplace „trebuie situat pe primul loc între acei savanți care, într-un fel, au aruncat o punte între aritmetica politică și calculul probabilităților" [95, p. 13]. La aceeași epocă, Gauss [26] definește legea normală de repartiție, pusă în lumină încă de A. de Moivre, Fourier [27] publică regulat considerații despre aplicarea matematicii în investigațiile demografice, iar Poisson [28] descoperă legea de repartiție în cazul evenimentelor rare și se ocupă de probleme de demografie, de statistică morală și judiciară, cu aplicații ale calculului probabilităților.

Numeroși alți matematicieni și învățați au realizat, în sec. al 19-lea, lucrări valoroase, care au consolidat gîndirea probabilistă în statistică și i-au asigurat progresul. Cel care a dominat însă ca statistician epoca sa a fost Quételet [29]. Cu el s-a produs o nouă cotitură în evoluția conceptului de statistică.

De obicei, numele lui Quételet este asociat cu teoria „omului mediu", un *homo statisticus*, care venea să înlocuiască pe *homo oeconomicus* al clasicilor. Conform acestei teorii statistice, de la care se pleca în constituirea unui fel de

[24] Jacob Bernoulli (1654—1705), profesor de matematică la Universitatea din Basel, a refăcut și extins lucrarea *De ratiociniis in ludo aleae*, prima carte de teoria probabilităților, scrisă de aritmeticianul politic olandez Christian Huyghens (1629—1695). Știința probabilităților a fost mult mai aprofundată de Bernoulli, în lucrarea *Ars conjectandi* (apărută la 7 ani după moartea sa), lucrare „încă remarcabilă prin justețea și finețea vederilor" (Laplace).

[25] Pierre-Simon Laplace (1749—1827), matematician, fizician și astronom francez, a efectuat în 1801, în Franța, un recensămînt pe bază de selecție și a calculat și coeficientul de eroare [80]. Lucrarea sa *Théorie analytique des probabilités* „va face în final să se convingă statisticienii de profitul ce-l vor putea trage din noua disciplină" [46].

[26] Karl Friedrich Gauss (1775—1855), matematician, fizician și astronom german, a propus cunoscuta metodă a celor mai mici pătrate, în lucrarea, scrisă în latină, *Theoria motus* (1809).

[27] Jean Baptiste Joseph Fourier (1768—1830), matematician francez, a promovat aplicațiile matematicii în statistică prin studii publicate în „Recherches statistiques sur la ville de Paris et le departement de la Seine".

[28] Simon Denis Poisson (1781—1840), matematician francez, a generalizat teorema lui Bernoulli cu privire la legea numerelor mari. Însăși denumirea de lege a numerelor mari *(loi de grands nombres)* a fost folosită prima oară de Poisson.

[29] Adolphe Quételet (1796—1874), matematician, statistician și astronom belgian, a extins cîmpul de aplicare al statisticii la studiul caracterelor fizice și morale ale oamenilor. În repetate rînduri s-a deplasat la Paris, unde a luat contact cu matematicieni francezi, în particular cu Laplace și Fourier, care l-au inițiat în lucrările lor referitoare la calculul probabilităților. În amintirile sale, povestește că s-a gîndit, la întoarcerea de la Paris, să aplice statistica la „dimensiunile corpului uman", pentru a-și aprofunda studiile de sculptură. Sinteza unor studii asupra statisticii o face în lucrarea *Sur l'homme et le développement de ses facultés ou essai de physique sociale* (1835), cuprinzînd elemente de statistică a populației, statistică morală și antropometrie, pe lîngă generalități despre organizarea sistemului social. Quételet a reluat ideile și îndemnul lui Laplace, cum arată însuși motto-ul lucrării: „Să aplicăm în științele politice și morale metoda întemeiată pe observație și pe calcul, metodă care ne-a servit atît de bine în științele naturale" *(Essai philosofique sur les probabilités)*. *Fizica socială* a fost socotită de o generație de statisticieni drept carte de căpătîi.

Scanned with OKEN Scanner

fizică socială, diversele caractere fizice, intelectuale și morale ale oamenilor, observați în masă, s-ar echilibra într-o ființă fictivă, ideală, dorită de natură, dar neizbutită; oamenii „reali" s-ar distribui în jurul ei „la întîmplare", analog erorilor în măsurarea unei mărimi. În lucrarea *Essai de physique sociale* citim: „Noi trebuie, înainte de toate, să pierdem din vedere omul luat izolat și să nu-l considerăm decît ca o fracțiune a speciei. Despuindu-l de individualitatea sa, vom elimina tot ceea ce nu este decît accidental, iar particularitățile individuale, care nu exercită decît puțină sau nici o acțiune asupra masei, se vor șterge și vor permite sesizarea rezultatelor generale" [110, p. 4—5].

Lăsînd la o parte latura sa fabuloasă, conceptul „om mediu" se cuvine evocat pentru ideile implicate: repartiție, medie, dispersie, observare de masă, regularitate, esențiale în cercetările statistice. „Cînd Quételet — scrie Durkheim — semnala atenției filozofilor surprinzătoarea regularitate cu care anumite fenomene sociale se repetă, el ținea seama de teoria sa despre omul mediu, care a rămas, dealtfel, singura explicație sistematică a acestei remarcabile proprietăți" (în lucrările sale — *n.n.*) [30].

Concepută ca o „fizică socială", statistica quételetiană încorporează ideea că regularitățile descoperite prin observarea de masă rezultă din legi care se impun omului în ordinea fizică și de asemenea în ordinea morală. „Eu nu am încetat o clipă să repet — declara Quételet cu referire la criminalitate — că există un buget ce se plătește cu o regularitate nemaipomenită, acesta este bugetul închisorilor, ocnelor și eșafodurilor...; anual numerele confirmă previziunile mele pînă într-un asemenea grad, încît aș putea să spun cumva cu mai multă exactitate: există un tribut pe care omul îl plătește cu mai mare regularitate decît tributul datorat naturii sau vistieriei statului — este tributul pe care-l plătește crimei" [110, p. 9]. Acest punct de vedere determinist-mecanicist era de nuanță moderată, deoarece admitea posibilitatea transformării ambianței sociale. Astfel, după ce afirmă că „societatea conține în sine germenii tuturor crimelor ce se vor comite" și că „orice stare socială presupune deci un anumit număr de crime și o anumită categorie de delicte, care apar drept consecință a organizării sale", Quételet continuă: „Această observație, care poate părea descurajatoare din primul moment, devine consolantă dimpotrivă cînd se examinează cu atenție, fiindcă ea arată posibilitatea de a îndrepta oamenii, modificînd instituțiile lor, obiceiurile lor, nivelul cunoștințelor lor și, în general, tot ceea ce influențează felul lor de a fi" [110, p. 11].

O altă lucrare a lui Quételet, cu semnificație aparte în dezvoltarea istorică a conceptului de statistică, este manualul *La méthode statistique* (1848) scris, în calitate de tutore, pentru prințul Albert al Belgiei. Elementele de statistică formulate, fără preocupare de sistematizare, în *Essai de physique sociale* și alte lucrări, au fost dezvoltate și întrunite, cu prilejul manualului, într-o expunere coerentă. Avem de-a face, pentru prima oară, cu o teorie sintetică a statisticii, cu un corp distinct de principii, tehnici și procedee statistice.

[30] Citat după [52, p. 3]. Teoria lui Quételet despre omul mediu a fost ținta unor atacuri foarte vii, fapt de altminteri anticipat de el însuși, cînd, în prefața la eseul său de fizică socială, și-a exprimat teama „de a fi rău înțeles în expunerea unei teorii, care va putea ofensa profund opiniile" [110, p. I]. Unii dintre criticii săi au socotit chiar că-i fac un serviciu, debarasîndu-i opera de o teorie discreditată și care nu i-ar fi deloc caracteristică. „Omul mediu al lui Quételet a murit înaintea autorului său ... Nimeni, din zilele noastre, nu se ocupă de el" (J. Lottin). Vezi [52, p. 2].

Scanned with OKEN Scanner

Elaborat după o perioadă de ample cercetări întemeiate pe practica statistică [31], manualul *La méthode statistique* corespunde unui stadiu mult mai evoluat al gîndirii autorului său. E adevărat, în primele sale lucrări, Quételet nu apare cu o opinie despre statistică prea diferită de opinia curentă. Statistica este numită „anatomia societății" și i se atribuie scopul „de a observa modificările pe care le suferă diferitele popoare în starea lor fizică și morală și de a căuta să pătrundă motivele" [61, p. 25]. Treptat însă, conceptul de statistică include elemente noi, îndreptînd statistica pe căi noi: cercetarea variației și a relațiilor cauzale în masa fenomenelor, dezvăluirea regularităților, investigarea cu predilecție a fenomenelor vieții umane etc. În final, Quételet ajunge să privească statistica drept singura metodă utilizabilă pentru studiul fenomenelor de masă „în care nu sîntem în stare să facem să varieze circumstanțele" și care sînt susceptibile „de a fi tratate prin calculul probabilităților"*.

1.1.3 Alte concepții post-quételetiene

Interpretarea probabilistă a statisticii, caracteristică lui Quételet, este comună și altor oameni de știință. După A. Cournot, unul din fondatorii economiei matematice, statistica ar fi „știința care are ca obiect culegerea și coordonarea faptelor numeroase de orice fel, încît să se obțină raporturi numerice sensibil independente de anomaliile hazardului și care denotă existența cauzelor regulate a căror acțiune s-a combinat cu cea a cauzelor întîmplătoare" [32].

Noua orientare n-a întîrziat să pătrundă în gîndirea statistică a epocii, dar controversele în jurul conceptului de statistică au continuat; definițiile în circulație s-au înmulțit [33]. Cel mai mult controversat era domeniul atribuit statisticii: unii admiteau, în sfera de cuprindere a statisticii, toate fenomenele exprimabile numeric, indiferent de natura lor [34], alții, numai fenomenele sociale [35], iar un alt grup restrînge cîmpul de aplicare al statisticii la fenomenele vieții umane („societatea civilă") [36]. Multă vreme încă, în secolul al 19-lea, autori de talent au păstrat, cu unele nuanțe, conceptul de statistică descrip-

[31] O vreme, Quételet a practicat curent statistica la biroul statistic al Țărilor de jos, iar, după despărțirea celor două țări, la biroul statistic belgian, pe care l-a condus. Mediul intelectual în care și-a format concepția despre statistică era impregnat de ideile filozofice și sociale caracteristice secolului luminilor. Împreună cu demograful și statisticianul englez William Farr, A. Quételet a organizat primul Congres internațional de statistică (Bruxelles, 1853), urmat apoi de alte opt congrese, care au jucat un important rol în dezvoltarea statisticii și a comparabilității sale internaționale.

* ANDRÉ LIESSE, *La statistique — ses procédés, ses résultats*, Paris, 1905, p. 7.

[32] Augustin Antoine Cournot (1801—1877), matematician și filozof francez, poate fi considerat unul din precursorii econometriei. Concepția sa despre statistică, sintetizată în definiția reprodusă, o găsim mai ales în lucrarea *Exposition de la théorie des chances et des probabilités* (1843).

[33] Gustav v. Rümelin (1815—1889), statistician german, a numărat 63 de definiții (inclusiv pe a sa) în 1863.

[34] Moreau de Jonnès: „Statistica este știința faptelor naturale, sociale și politice, exprimate în termeni numerici" [93, p. 1].

[35] P.A. Dufau: „Știința care ne învață să deducem din termeni analogi legile succesiunii faptelor sociale" [41, p. 35]. Ca știință a faptelor sociale a fost definită statistica și de Moreau de Jonnès, în prima ediție a lucrării sale *Éléments de statistique* (1847).

[36] Wilhelm Lexis: „Statistica îndeplinește funcția proprie de a culege și studia, după o metodă exactă, faptele vieții umane, luate în masă..." [61, p. 32].

tivă al şcolii din Göttingen sau conceptul de aritmetică politică la faza iniţială şi au scris lucrări în concordanţă cu conceptul însuşit [37].

O remarcabilă contribuţie la progresul statisticii, al afirmării ei ca ştiinţă au adus şi alte şcoli naţionale, în decursul sec. al 19-lea.

Şcoala germană a continuat, în linii generale, tradiţia unei statistici, concepută ca ştiinţă socială, cu numeroase elemente de statistică descriptivă. Unul din reprezentanţii de seamă este Wilhelm Lexis [38], după care domeniul statisticii îl constituie societatea, procesele care au loc în cadrul ei, ca procese de masă, supuse legilor statistice. Aproximativ pe aceeaşi linie se situează şi statisticienii demografi germani Karl Becker (1823—1896) şi mai ales Georg von Mayr (1841—1925). Pentru acesta din urmă, statistica era o ştiinţă socială generală, un fel de sociologie *(Statistik und Gesellschaftslehre)*. Istoricii statisticii apreciază că, datorită puternicii autorităţi a lui G. v. Mayr, orientarea cantitativă şi, în general, statistica matematică au pătruns greu în statistica germană.

Statistica rusă a fost reprezentată, în sec. al 19-lea, atît în statistica socială, cît mai ales, în cea de-a doua jumătate a secolului, în teoria probabilităţilor. O figură originală este D. P. Juravski [39], după care statistica este o ştiinţă a „categoriilor", a „comparaţiei", punînd un mare accent pe metoda grupărilor, o ştiinţă materială avînd ca obiect studiul vieţii sociale şi, în acelaşi timp, ştiinţă metodologică. La teoria probabilităţilor, contribuţii fundamentale au adus P. L. Cebîşev (1821—1894) şi A. A. Markov (1866—1922). Lui Cebîşev se datorează, printre altele, demonstrarea teoremei centrale a teoriei probabilităţilor, a legii numerelor mari, iar Markov a pus bazele teoriei moderne a proceselor stohastice şi a lanţurilor care-i poartă numele. De asemenea, trebuie amintit A. A. Ciuprov (1874—1926), care, primul, a încercat să facă o sinteză a curentului german reprezentat de W. Lexis şi L. v. Bortkiewicz şi a curentului anglo-saxon, afirmat la sfîrşitul sec. al 19-lea prin F. Galton şi K. Pearson.

Desigur, menţiune merită şi statistica din alte ţări, cum ar fi Italia, cu o veche tradiţie, reprezentată în sec. al 19-lea şi începutul sec. al 20-lea, de nume ca Luigi Bodio (1840—1920) şi Corrado Gini (1884—1965).

1.1.4 Gîndirea statistică românească în secolul al XIX-lea

În ansamblul curentelor statistice ale timpului s-a încadrat organic gîndirea statistică românească. Pionierii statisticii româneşti au avut un strîns contact cu realizările din alte ţări europene şi au adus contribuţii remarcabile la pro-

[37] Gioja Melchiorre: „În acest sens (sensul cuvîntului italian *stato, n.n.*) statistica înseamnă descrierea calităţilor care caracterizează sau a elementelor care compun un stat. Statistica arată: bogăţia sau sărăcia, ştiinţa sau ignoranţa, fericirea sau nefericirea, moralitatea sau corupţia, civilizaţia sau barbaria, puterea sau slăbiciunea naţiunii". Melchiorre distinge apoi *filozofia statisticii*, definită astfel: „...cunoaşterea raţională a normelor generale de cercetare, a surselor de date, a criteriilor de verificare, a principiilor de a judeca, a aplicaţiilor la elementele care privesc starea naţiunilor" [48, p. 1 şi 7].

[38] Wilhelm Lexis (1837—1914), statistician, economist şi demograf, şi-a expus teoria mai ales în *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft* (1877). Este unul din fondatorii „statisticii morale". A introdus teoria „stabilităţii seriilor statistice", criteriul de divergenţă, ce-i poartă numele, pentru aprecierea diferitelor tipuri de variaţii. Şi-a cîştigat merite deosebite în demografie.

[39] D. P. JURAVSKI (1810—1856), în special cu lucrarea sa *Ob istocinikah i upotreblenii statisticeskih svedenii* (1846 şi reeditată în 1946).

Scanned with OKEN Scanner

movarea statisticii, atît pe planul teoretic, cît şi al organizării ei instituţionale [40]. Secolul al 19-lea, denumit sugestiv „era entuziasmului statistic“ [41] (H. Westergaard), este marcat în ţara noastră de personalităţi proeminente ale cugetării statistice: Ion Ionescu de la Brad şi Dionisie Pop Marţian.

Ion Ionescu de la Brad [42] şi-a expus concepţia despre statistică în lucrarea *Povăţuiri pentru catagrafia Moldovei* (1859), un sintetic manual de teorie a statisticii. Din numeroasele definiţii date statisticii şi care îi erau familiare, el a reţinut, după o filtrare atentă, definiţia: „Statistica este ştiinţa faptelor sociale exprimate prin termeni numerici“ (Moreau de Jonnès), socotind că reflectă „adevărata natură a statisticii“, am spune caracterul ei de ştiinţă socială. Împotriva teoreticienilor care „lărgeau în nemărginire hotarele statisticii“ — şi se referea la primii „apostoli“ ai statisticii —, Ionescu de la Brad a subliniat „grija de a nu despărţi ideile amîndouă ce se cuprind în formulă: ideea faptelor sociale de aceea a faptelor arătate prin cifre“ [59, p. 1—2].

Elocvente pentru gîndirea statistică înaintată a lui Ionescu de la Brad sînt precizările sale de ordin metodologic. Statistica — arată el — se alcătuieşte din două părţi: „statistica propriu-zisă“ — culegerea, coordonarea şi eventual reducerea datelor şi „partea curat matematică“ — conjecturarea faptelor necunoscute prin faptele cunoscute şi prin „artificiile socotelilor“. De unde decurg două metode de a dobîndi informaţii statistice despre faptele sociale: „metoda naturală sau expozitivă“, specifică statisticii descriptive, şi „metoda de inducţie“, specifică aritmeticii politice [59, p. 3, 5—6]. Recunoaştem în aceste modalităţi metodologice, în linii mari, însuşi demersul statistic modern în două etape: una descriptivă, alta inductivă.

După Dionisie Pop Marţian [43], statistica studiază, în expresie numerică, „fenomenele naturale şi sociale“ şi „are de scop a conduce la descoperirile cauzelor şi efectelor şi de aici la legile care domnesc viaţa universală a naturii şi cea particulară a popoarelor şi a arăta prin coordonarea acestor legi unitatea şi armonia lumii fizice şi sociale în deosebitele lor fenomene“ [75]. Rezultă în mod evident adeziunea sa la teza de largă circulaţie potrivit căreia obiectul de cercetare al statisticii excede sfera fenomenelor sociale. Totuşi, în tot ceea

[40] Un prilej de cunoaştere a concepţiilor şi înfăptuirilor statistice din acea vreme l-au constituit congresele internaţionale de statistică, care au reunit reprezentanţi ai statisticii oficiale şi ştiinţifice din diferite ţări, comunicînd experienţele şi vederile lor în domeniul practicii şi teoriei statistice. România a participat la patru dintre cele nouă congrese ţinute între anii 1853 şi 1877.

[41] Caracterizat de înfiinţarea oficiilor naţionale de statistică, de efectuarea primelor recensăminte moderne şi de apariţia societăţilor de statistică şi revistelor de statistică.

[42] Ion Ionescu de la Brad (1818—1891), agronom, economist şi statistician, a organizat şi a condus primul oficiu naţional de statistică din Moldova. Lucrarea *Povăţuiri pentru catagrafia Moldovei, precedate de oarecare elemente de statistică*, scrisă în legătură cu recensămîntul din 1859—1860 în Moldova, poartă caracterul unui adevărat manual de statistică, cel dintîi de acest gen apărut în limba română [11, p. 15]. Literatura statistică apuseană îi era bine cunoscută, după cum reiese clar din lucrare, mai cu seamă din timpul studiilor făcute în Franţa şi Belgia. Ion Ionescu de la Brad a reprezentat ţara la Congresul internaţional de statistică de la Haga (1869).

[43] Dionisie Pop Marţian (1829—1865), economist şi statistician, a organizat primul oficiu naţional de statistică din Ţara Românească şi apoi a condus organul central de statistică al Principatelor Unite. Din iniţiativa şi sub conducerea sa a apărut, în 1860, revista „Anale statistice“, prima publicaţie statistică românească. A reprezentat ţara la Congresul internaţional de statistică de la Berlin (1863).

Scanned with OKEN Scanner

ce a scris D. P. Marţian relativ la statistică predomină afirmarea rolului ei de ştiinţă socială, ca, de exemplu, în formularea: „... a se cunoaşte societatea pe sine, iată scopul şi tendinţa finală a statisticii" [75]. În altă parte, întocmind o „tabelă a ştiinţelor sociale", a înscris la grupa ştiinţelor sociale generale „istoria filozofică a societăţilor" şi „statistica socială", cu motivarea: „Cum s-a dezvoltat societatea spre a fi ceea ce este acum? — şi ceea ce este acum? La prima întrebare ne răspunde: istoria; la a doua: statistica" [76, p. 81].

În studiile sale, D. P. Marţian insistă asupra trăsăturii distinctive a statisticii — cercetarea realităţilor sub aspectul cantitativ: „... tot ce ştim şi înţelegem, ştim şi înţelegem (în statistică — *n.n.*) prin ajutorul numărului şi a măsurii" [77] şi susţine posibilitatea descoperirii cu ajutorul statisticii a legilor care guvernează procesele investigate. Asociind, în mod necesar, ideii de lege statistică ideea că statistica studiază fenomenele de masă din punct de vedere cantitativ, observăm cît de mult s-a apropiat D. P. Marţian de înţelegerea particularităţilor de seamă ale statisticii, azi consacrate.

Ca şi în alte ţări, s-au scris şi în ţara noastră, în secolul al 19-lea, lucrări statistice în spiritul statisticii descriptive înaintate (şcoala statisticii universitare) [44]. În această accepţie, statistica a figurat şi ca disciplină în învăţămîntul românesc. Un curs, „Statistica generală a Europei", a fost deschis în 1856, la Facultatea juridică de la Academia Mihăileană din Iaşi, de către Alexandru Papiu Ilarian, avînd ca obiect „staturile din care se compune Europa, şi anume toate condiţiunile care constituiesc puterea şi fericirea actuală a acestor staturi". Asupra acestei concepţii, declară Papiu Ilarian, „se învoiesc mai toţi cultivatorii ştiinţei noastre" [45].

1.1.5 Statistica modernă

Către sfîrşitul secolului trecut se conturează o nouă etapă în evoluţia procesului de conceptualizare a statisticii — etapa constituirii *statisticii moderne.*

Depăşirea stadiului descriptiv, al simplei reprezentări cantitative, a însemnat, istoric, deplasarea gîndirii statistice spre *interpretarea analitică* a fenomenelor de masă şi obţinerea de *concluzii inductive*, pe baza observaţiilor empirice. Această schimbare de conţinut a făcut ca statistica să se întrepătrundă în mod constant cu matematica, în căutare de metode corespunzătoare obiectivelor sale. O mare parte din metodele obişnuite de lucru ale

[44] Apartenenţa la curentul descriptivist în statistică este mărturisită deschis de economistul Nicolae Şuţu (1798—1871) în lucrarea *Études statistiques* (1837): „statistica... nu face altceva decît să descrie, iar încă o particularitate care ar trebui s-o distingă de alte ştiinţe de asemenea descriptive constă în aceea că faptele pe care ea le descrie sînt variabile" [119, p. 300]. Un alt reprezentant tipic al acestui curent a fost George Barițiu (1812—1893), care, situînd statistica alături de istorie, observă că sarcina lor principală este de a analiza mersul şi natura evenimentelor sociale, de „a cerceta cauzele şi piedicile" ce stau în calea dezvoltării istorice a societăţii omeneşti [10].

[45] Cu privire la conţinutul cursului, păstrat în manuscris, şi la concepţia despre statistică a lui Al. Papiu Ilarian (1828—1878), a se vedea [33]. În descrierea statelor, Papiu a folosit metoda comparativă, care i-a permis „să sublinieze trăsăturile comune ale organizării politice româneşti cu alte state europene, dar mai ales să pună în lumină originalitatea instituţiilor româneşti" [25].

Scanned with OKEN Scanner

analizei statistice au fost elaborate de școala anglo-saxonă de statistică matematică, fondată de F. Galton și K. Pearson și continuată de R. A. Fisher și alții [46].

Dintru început statistica se îmbogățește cu ideea de „variație legată" a fenomenelor, dezvăluită de F. Galton [47]. Transpusă în limbaj matematic de către K. Pearson [48], ea s-a constituit apoi într-o teorie vie — teoria corelației statistice —, ale cărei aplicații au permis identificarea unor „noi forme de legi de dependență, specific statistice și admirabil adaptate la formele complexe și variate pe care le oferă viața în diferitele sale manifestări" [106].

R. A. Fisher [49] preconizează extinderea preocupărilor statistice la trei categorii de probleme: a) probleme de specificație — expresia matematică a populației; b) probleme de estimație — depistarea metodelor convenabile pentru calculul coeficienților proprii estimației parametrilor necunoscuți ai populației; c) probleme de distribuție — distribuția aleatoare a estimațiilor parametrice și a altor coeficienți aleși pentru a testa validitatea specificației. Probabiliștii și statisticienii sec. al 19-lea au avut în vedere numai prima categorie de probleme. În demersul modern se pornește de la *ipoteze statistice* asupra fenomenului sau procesului observat, consecințele lor logic deduse se compară apoi cu datele disponibile și dacă sînt în concordanță, ipotezele sînt justificate, cel puțin pînă la noi observații mai rigurcase.

În accepția modernă, caracteristică pentru statistică este preocuparea de a trage, printr-o analiză subtilă, concluzii de o valoare generală din informații culese prin cercetări pe eșantion. Dacă mai înainte aproape întreaga activitate statistică se sprijinea pe *observarea totală*, acum *sondajul* tinde să devină de aplicație curentă. Dar a conchide de la eșantion la populație presupune o anumită incertitudine, pe care cercetătorul caută să o reducă. Unul din obiectivele esențiale ale statisticii, care-i dă nota sa originală, se consideră tocmai măsurarea incertitudinii concluziilor inductive. De la un timp însă, gîndirea statistică tinde să se preocupe mai puțin de măsurarea incertitudinii și mai mult de determinarea riscurilor de eroare și a pierderilor implicate de orice decizie întemeiată pe o informație care, prin natura sa, nu poate fi exhaustivă.

Sub impulsul școlii Pearson-Fisher, statistica a făcut progrese remarcabile, găsindu-și fundamentări foarte precise ca disciplină științifică. Expansiunea cercetării științifice, îndeosebi în domeniul teoriei matematice a statisticii, în diferite țări, pe căi originale, a dus la descoperirea de procedee

[46] Statistica matematică, bazată pe teoria probabilităților, din care s-a desprins, și cultivată mai întîi în Germania (W. Lexis) și Rusia (A.A. Ciuprov), a devenit în special un domeniu al statisticienilor englezi. Poziția proeminentă a școlii engleze se datorează puternicilor personalități care au fost Karl Pearson și Ronald A. Fisher. Printre numele de prestigiu din spațiul anglo-saxon se numără F.Y. Edgeworth, G.U. Yule, M.G. Kendall. Contribuțiile lor sînt hotărîtoare pentru formarea statisticii moderne, elaborînd sau dezvoltînd capitole de bază ca: analiza dispersională, calculul de corelație, teoria estimației, verificarea ipotezelor, teoria selecției.

[47] Francis Galton (1822—1911), biolog, antropolog și statistician, a pus bazele biometriei. În 1901 a întemeiat revista „Biometrika", care a avut un rol de seamă în dezvoltarea statisticii matematice.

[48] Karl Pearson (1857—1936), statistician, antropolog, psiholog și filozof, a adus contribuții însemnate la teoria corelației, a distribuțiilor („curbele Pearson"), a momentelor de repartiție. Cele mai multe studii de statistică le-a publicat în „Biometrika".

[49] Ronald Aylmer Fisher (1890—1962), matematician, statistician și biolog, a condus Stațiunea experimentală de la Rothamsted, un adevărat laborator în care s-au făurit diferite metode ale statisticii matematice și s-au aplicat la experimente, mai ales de agrobiologie. Lucrarea principală: *Statistical Methods for Research Workers* (1925) a cunoscut numeroase ediții.

Scanned with OKEN Scanner

de investigație din ce în ce mai elaborate, mai eficiente și la cucerirea unor noi sfere de aplicație statistică [50]. În ultimele decenii, statistica a pătruns, ca metodă științifică, în cele mai variate zone ale științelor naturale și sociale. Cu arsenalul de procedee de care dispune acum, se poate spune că ea reușește pretutindeni unde trebuie să se studieze fenomene de masă, transcrise sub formă de date numerice.

1.1.6 Cîteva concluzii

Incursiunea istorică îngăduie precizarea surselor statisticii moderne și demonstrează ceea ce nu se mai înțelege și ceea ce totuși continuă să se înțeleagă prin statistică.

Conceptul de statistică are, după cum s-a văzut, o îndelungată istorie. Deși nu a fost formulat expres decît mult mai tîrziu, premisele sale au fost puse încă din antichitate; rădăcinile teoretice le aflăm în *statistica descriptivă*, *aritmetica politică și teoria probabilităților*.

De-a lungul celor trei secole de existență ca disciplină științifică, statistica a fost definită foarte diferit, fapt ce se explică, pe de o parte, prin împrejurările istorice în care a apărut și s-a dezvoltat, iar, pe de altă parte, prin extinderea considerabilă a cîmpului de aplicare a metodelor statistice. Dacă dăm crezare lui Walter Willcox [51], numărul definițiilor date conceptului de statistică de către diverși autori, între 1749 și 1934, a depășit o sută. Această „inflație" a continuat deoarece „printre temele în legătură cu care statisticienii nu sînt de acord, se găsește definiția științei lor" (Kendall). O clasificare a definițiilor înregistrate este dificilă și ar fi, credem, în bună măsură artificială, întrucît adeseori se amestecă, de fapt, puncte de vedere diferite [52].

În diversitatea definițiilor date statisticii, se disting două opinii: unii califică statistica drept *știință*, alții o admit numai ca *metodă*. Dar, într-o înțelegere dialectică, nu este vorba de laturi antitetice, căci „ca orice știință, statistica își are metoda sa proprie de cercetare, determinată atît de obiectul său propriu — fenomenele și procesele sociale de masă —, cît și de baza sa

[50] Jerzy Neyman, american de origine polonă, împreună cu Egon S. Pearson, fiul lui K. Pearson, au introdus în teoria estimațiilor, inițiată de R.A. Fisher, noțiunile de riscuri de ordinul întîi și al doilea și noțiunea de puterea unui test și au construit teoria sondajelor. Student (pseudonimul lui W. Gosset) a adus contribuții importante la teoria eșantioanelor de volum redus, la legea de repartiție a mediilor, teoria erorii probabile a mediei, repartiția t (Student). Abraham Wald (absolvent al facultății de științe din Cluj-Napoca, în anul 1924) a elaborat teoria testului secvențial. Georg Udny Yule, colaborator al lui K. Pearson, a dezvoltat teoria corelației și teoria caracteristicilor calitative. Ragnar Frisch, economist norvegian, a propus o metodă de analiză a multicoliniarității în cazul regresiei multiple liniare (metoda fasciilor), iar statisticianul sovietic V. S. Nemcinov s-a ocupat de sistemele de eliminare succesivă a variabilelor independente la formularea modelelor de corelație și a întocmit o schemă proprie de rezolvare a sistemului de ecuații de regresie prin polinoamele lui Cebîșev („schema N").

[51] Walter F. Willcox (1861—1964), statistician și demograf american, autor al unor studii clasice asupra migrației internaționale și al estimațiilor populației globului, estimații ce-i poartă numele.

[52] Ca, de exemplu, clasificarea în: definiții umoristice și critice, definiții care confundă instrumentul cu domeniul de aplicare, definiții limitate [109, p. 59—62].

Scanned with OKEN Scanner

teoretică“ [71, p. 203]. Fiecare disciplină își merită numele prin aceea că elaborează și dispune de un corpus separat de cunoștințe, de tehnici și metodici. „Sub hainele variate ale feluritelor aspecte ale statisticii sînt un număr de elemente comune, care constituie schema științei statistice. ... Cu valori tipice, cu frecvențe, cu fluctuații, cu corelații și cu indici, cu cîteva scheme tipice, statistica speră să dea imaginea lumii, făcînd să renască pentru omul de știință iluziile pe care le-a trăit mecanicismul veacului trecut“ [O. Onicescu, în 47, p. X]. Statistica se dezvoltă azi ca o știință cvasiindependentă; ea are însușirea — comună științelor — de a fi, în același timp, știință și metodă utilizabilă și în alte științe.

Statistica a apărut ca o generalizare *sui-generis* a unor date despre fenomene sociale. Istoricește ea s-a situat în mediul științelor sociale. Încă pe la mijlocul sec. al 19-lea, sfera statisticii era considerată numai viața socială; mai pe urmă a început să se „infiltreze“ în domenii ale științelor naturii. În genere, adepții ambelor școli istorice — statistica descriptivă și aritmetica politică — au optat pentru caracterul social al statisticii și mai consecvent în Germania. La începutul perioadei moderne a statisticii, Georg von Mayr [53], revendicînd energic pentru statistică dreptul la demnitatea de știință, a definit-o „știința maselor sociale de orice fel, în măsura în care studiul lor se efectuează prin metoda analizei numerice“ *.

Alături însă de statistica socială s-au format și s-au dezvoltat, în decursul timpului, noi discipline statistice: statistica matematică, statistica fizică, statistica biologică etc. Astăzi, statistica are numeroase ramuri, care au ca obiect fenomene sociale și nesociale; vocația ei este largă. Accepțiile curente ale noțiunii de statistică sînt și ele multiple: *date statistice* — date cantitative afectate în mare măsură de o mulțime de cauze; *metode statistice* — metode special adaptate la studierea datelor statistice; *metodă statistică* — metodă științifică de tip special care se aplică în procesul de cunoaștere a fenomenelor de masă; *teoria statisticii* sau *statistică* — disciplină care se ocupă cu expunerea sistematică a conceptelor, principiilor și metodelor de cercetare statistică [124, p. 41 și 134, p. 14 și 16].

Concepută ca știință sau ca metodă, statistica modernă prezintă — după consensul general — două aspecte diferite și, de altminteri, complementare: *descrierea statistică* — fixarea informației așa cum a rezultat din prelucrarea datelor empirice și *inferența statistică* — tratarea teoretică a datelor pentru a trage concluziile logice, asociate observațiilor efectuate. Primul aspect, mai puțin ambițios, evocă partea clasică a statisticii și este în general cunoscut sub numele de *statistică descriptivă*, iar al doilea aspect reprezintă *statistica inductivă*. În demersul statistic, cele două aspecte trebuie să fie privite ca două momente ale unuia și aceluiași proces de cunoaștere [54].

[53] GEORG VON MAYR, *Statistik und Gesellschaftslehre* (3 vol., 1895—1917).

* CAMILLE JACQUART, *Statistique et science sociale*, Bruxelles, 1907, p. 37—38.

[54] După concepția frecvent întîlnită în statistica franceză, treptele ar fi: statistica descriptivă — teoria probabilităților — statistica matematică, legătura dintre statistica descriptivă și cea matematică fiind asigurată de conceptul de „probabilitate“. În mod corespunzător, în felul acesta se realizează trecerea de la „repartiția empirică“ la „repartiția teoretică“. A se vedea G. CALOT, *Cours de statistique descriptive*, Dunod, Paris, 1965.

Scanned with OKEN Scanner

1.2 FENOMENELE DE MASĂ CA OBIECT DE INVESTIGAȚIE STATISTICĂ

Puternic controversată, supralicitată de unii, defăimată de alții, statistica se dovedește indispensabilă cunoașterii *fenomenelor de masă*, generate de cauze multiple și complexe. Rațiunea de a fi a statisticii constă în studiul cantitativ al acestor fenomene, care, datorită întinderii, diversității și variabilității lor, nu apar sau nu apar cu exactitatea dorită decît la nivelul mulțimii sau colecției de observații [55].

Un fenomen dat se poate prezenta în formă simplă, cu un număr restrîns de cauze cunoscute și dincolo de care este de prisos să se caute altele, cînd o anume aproximație se consideră satisfăcătoare. În consecință, fenomenul nu-i greu de apreciat prin măsuri precise; pentru studierea lui, observația și experimentul sînt metodele de predilecție. Alte fenomene, luate în studiu, decurg dintr-o împletire inextricabilă de cauze, principale și secundare, esențiale și neesențiale, de natură să imprime o mare variabilitate a cazurilor singulare. Asemenea fenomene scapă perceperii directe și individuale, impunînd considerarea în masă și tratarea lor statistică.

Fenomenele de masă se întîlnesc aproape în toate domeniile științei și ale activității sociale. Sub aspectele cele mai complexe apar ele însă în domeniul social. În fenomenele sociale — obiectul primordial al statisticii — se ivesc forme de relații cauzale deosebit de înalte: afară de cauzele esențiale și neesențiale intervin lanțurile de evenimente cauzale, în care se integrează, ca o verigă, scopurile umane — personale și sociale. Fenomenalitatea neobișnuit de complexă a vieții sociale ridică cele mai mari dificultăți în stabilirea schemelor fenomene-cauze supuse analizei statistice.

Aplicarea calculelor statistice la datele empirice, oferite de observarea fenomenelor de masă, permite desprinderea *legilor statistice*. Aceste legi descriu relații cauzale statistice. Ele sînt legi cu statut propriu, tot așa de obiective ca și legile dinamice.

După tipul relațiilor fixate, legile se împart în *legi dinamice* și *legi statistice*. Legile dinamice acționează în fiecare caz singular în parte, fixînd tocmai generalul necesar care se manifestă în toate cazurile singulare. În circumstanța acestor legi, prezența factorilor întîmplători se face simțită sub forma abaterilor statistice de la desfășurarea pură a necesității. Modelul legilor dinamice îl reprezintă legile mișcării mecanice (de exemplu: legea căderii corpurilor, legea lui Arhimede etc.). Legile statistice exprimă *media* stărilor unei mase de evenimente; în cadrul ei se resimte mai puternic acțiunea factorilor întîmplători (*Materialism dialectic*, Editura politică, București, 1973, p. 263).

Particularitatea esențială a legilor statistice izvorîtă din faptul că ele acționează în fenomenele de masă, unde întregul este determinat de unitatea părților componente, o constituie exprimarea comportării ansamblului de unități omogene, și nu a fiecărei unități în parte. În mod corespunzător, legea statistică se realizează „ca o tendință predominantă, ca o necesitate care își croiește drum printr-un număr foarte mare de contingențe și care se manifestă

[55] Conceptul de *fenomen* se folosește în materialismul dialectic, ca și în științele speciale, pentru denumirea unei acțiuni comune a numeroase elemente omogene, a unui ansamblu de elemente luate ca întreg în legătura lor comună, și nu pentru denumirea unui obiect individual. O infracțiune, de exemplu, este un eveniment, în timp ce infracționalitatea este un fenomen. Demografia oferă cazuri similare: o naștere este un eveniment statistic; frecvența lui în populație, expusă riscului de naștere, este natalitatea sau fertilitatea populației respective.

Scanned with OKEN Scanner

în aceste contingențe ca media unui număr mare de abateri întîmplătoare" *. De aici și principiul verificat adeseori de practică: legea statistică poate fi evidențiată dacă și numai dacă este considerat și supus observării un număr mare sau suficient de mare de unități elementare ale ansamblului studiat.

Cei dintîi statisticieni, care au constatat regularități în fenomenele sociale cercetate, s-au referit statornic la principiul numerelor mari. J. P. Süssmilch, de exemplu, a scris că *ordinea* în frecvența deceselor, nașterilor, căsătoriilor este vizibilă în numărul mare de cazuri: „o proprietate a acestei ordini constă în aceea că se ascunde în dezordinea numărului mic de cazuri și poate fi descoperită numai dintr-un mare număr de registre despre multe localități și pentru mulți ani" [56]. A. Quételet, căruia după expresia plastică a unui critic „îi tremura mîna de emoția bucuriei" consemnînd rezultatele investigațiilor sale statistice, a condiționat „cunoașterea legilor generale" de reunirea observațiilor „în destul de mare număr" pentru a elimina din ele „tot ceea ce nu este decît întîmplător și individual".

În colective mari, cuprinzătoare, are joc liber „legea numerelor mari", expresie a legăturii dialectice dintre necesitate și întîmplare. În virtutea acestei legi, acțiunea comună a unui mare număr de factori aleatori, în anumite condiții, duce la rezultate care aproape că nu depind de întîmplare, dînd deci posibilitate necesității să se realizeze. Ca ilustrare, să facem apel la prima lege statistică descoperită: la naștere, numărul băieților întrece pe cel al fetelor; raportul inițiat admis era de B 107: F 100. Căutăm să probăm păstrarea raportului de masculinitate la naștere în zilele noastre. Este în afară de orice îndoială că, dacă ne-am opri asupra distribuției născuților pe sexe în sînul fiecărei familii în parte, ar fi imposibil să indicăm ce fel de raport tipic există, date fiind fluctuațiile sale destul de numeroase și importante pentru diferite familii. O cercetare, singură posibilă și singură valabilă într-o atare situație, trebuie întreprinsă pe mase îndeajuns de mari de nașteri. Într-adevăr, observarea cazurilor de nașteri, de exemplu pe cuprinsul țării noastre, constituie mărturia faptului că an de an se nasc ceva mai mulți băieți decît fete [57]; raportul general este de B 106 : F 100.

Legea numerelor mari are o mare însemnătate nu numai „ca o teză abstractă a științei"; ea joacă un rol de seamă și în activitatea practică a oamenilor. J. Bernoulli, care pentru prima dată a îmbrăcat această lege în formă matematică, a ținut să menționeze că procedeul sporirii numărului observațiilor în scopul de a ridica gradul de siguranță al judecății „nu este nou și nu este neobișnuit" și că „chiar omul simplu, după un instinct natural oarecare, el singur și fără nici un fel de instruire prealabilă (lucru de mirare) știe că, cu cît mai multe observații sînt luate în considerare, cu atît este mai mic pericolul de a nu se ajunge la țintă" **.

* *Dialectica marxistă și științele moderne*, vol. II, Editura politică, București, 1960, p. 40–41.

[56] Regularitățile statistice au avut darul de a emoționa adînc pe primii cercetători. J.P. Süssmilch a socotit că se află în fața unei „ordine divine", care reglementează fatal toate gesturile omului. În opoziție cu Süssmilch, A. Quételet a așezat interpretarea ideii de constanță statistică, fără deosebire de împrejurări, sub influența legilor naturii.

[57] Raportul de masculinitate la naștere a oscilat între 1,052 și 1,071 în seria de ani 1955–1973, reprezentînd o mărime medie de 1,061 (vezi *Anuarul demografic al Republicii Socialiste România 1974*, p. 146–147). La scară mondială, se admite acest raport cu o valoare tipică de 1,06.

** Citat de N. K. DRUJININ, *Osnovnie matematiko-statisticeskie metodi v economiceskih issledovaniah*, Ed. „Statistika", Moscova, 1968.

Scanned with OKEN Scanner

Investigația statistică presupune, prin definiție, considerarea fenomenelor în multiplicitatea și variabilitatea lor. Un ansamblu de elemente formează un fenomen de masă sau, ceea ce numim, o *populație statistică* [58], în măsura în care elementele componente sînt de aceeași natură, adică au, toate, o proprietate comună și se deosebesc unele de altele în raport cu aspectele sau valorile caracteristicilor luate în studiu. Populația cu care operează statistica trebuie să fie *global omogenă* — să includă doar elemente similare, aparținînd de aceeași „categorie" și *intern structurată* — elementele ei să poată fi ordonate potrivit unui sistem de clasificare.

1.3 DEMERSUL STATISTIC

Pentru dobîndirea de cunoștințe autentice despre fenomenele cu caracter de masă, statistica posedă un mod de investigație care-i este propriu și un complex de procedee specifice. Calea spre cunoaștere pornește de la observarea masei de fenomene empirice, materializată în strîngerea datelor de bază, originale și se continuă cu prelucrări și analize succesive care legitimează concluziile statistice.

Concret, *observarea* poartă asupra unei populații în absoluta totalitate a unităților constitutive sau asupra unui eșantion reprezentativ. Fenomenele sînt observate, respectînd anume reguli metodice și folosind anume tehnici, în mediul lor natural de manifestare. Gradul de complexitate a fenomenelor de masă, atît structurală cît și cauzală, nu îngăduie observatorului să experimenteze, să le reproducă după voie, ci îl obligă să le considere așa cum ele au loc, în afara controlului său. Această împrejurare explică și dificultățile întîmpinate la nivelul analizei și interpretării cauzale a fenomenelor de masă.

Datele brute ale observării nu au o semnificație simplă și imediată; se impun operații rafinate de prelucrare pentru a le aduce la o formă clară și utilizabilă. În prelucrarea inițială, operația de prim ordin este *gruparea* — dezmembrarea populației în grupe omogene de unități, conform cu valorile sau accepțiile discriminante ale uneia sau mai multor caracteristici. Repartițiile statistice, rezultate din grupări, se prezintă simplu în *tabele* și, dacă împrejurările o cer, se transpun în *grafice*.

Odată ce materialul statistic a fost „turnat în forme" convenabile, se caută parametrii care pot descrie cel mai bine *cîmpurile de variabilitate* după diferitele lor caractere. Obiectivul este atins prin calcule de valori relative, valori medii, dispersii, abateri și alte valori tipice. Astfel de „specificări" definesc felul de statistică practicat în mod obișnuit, fie că este vorba de elemente măsurabile, fie că este vorba de repartiția unei calități înăuntrul unei populații.

Un alt grup de preocupări vizează o formă de realitate superioară în care se integrează populația cercetată și se concretizează în prelucrarea aprofun-

[58] Populație statistică sau simplu *populație*, termen moștenit de la cel dintîi domeniu de investigație statistică: populațiile umane. Acest termen s-a generalizat, desemnînd orice mulțime de natura descrisă: persoane, evenimente, obiecte etc. Prin analogie, se numește *individ* fiecare din unitățile elementare ale unei populații. Un cuplu de termeni sinonim este de asemenea uzual: *colectivitatea statistică* și *unitatea statistică*.

Scanned with OKEN Scanner

dată a *legăturilor* dintre elementele populației. Procedeele ce se folosesc sînt multiple; de ele se ocupă teoria asocierii, regresiei și corelației.

În stadiul următor al cercetării, se procedează la *aprecierea critică* a valabilității parametrilor empirici. Firește, parametrii calculați sînt definiți și precizi (cu condiția aplicării unor procedee corecte) numai în ce privește colecția de unități observate; ei pot fi luați drept reflectări mai mult sau mai puțin imprecise, mai mult sau mai puțin deformate ale unor parametri într-o populație mai întinsă, reală sau ipotetică, cunoscută sau necunoscută. Pentru estimații dincolo de datele disponibile, se aplică metodele de inferență statistică, intrînd în acțiune, ca un instrument complementar, conceptul de probabilitate. Mai întîi, prin inducții dirijate și controlate, se desvăluie modelele teoretice (probabiliste) aplicabile observațiilor; apoi, reprezentările concrete sînt confruntate cu modelele induse și se formulează — în termeni de probabilitate — gradul de încredere ce se poate acorda concluziilor. Și, în sfîrșit, acestui stadiu îi succed, dacă este cazul, *calcule de previziune.*

Cea mai mare parte a cercetărilor statistice — chiar cînd par exhaustive — sînt asimilabile „sondajelor" într-o populație infinită și impun parcurgerea completă a drumului metodologic schițat. Trebuie să admitem că, dacă nu s-ar realiza saltul spre elaborarea probabilităților, statistica nu s-ar libera niciciînd de cazurile particulare și vederile sale ar fi extrem de limitate. Probabilitatea „imprimă marca de adevărat obiect al științei statistice. Cei ce nu vor să se uite la această marcă pot, totuși, să facă știință bună dacă marca există. Dar ei nu o știu sau, mai curînd, nu vor să o știe. Dacă, însă, această marcă nu există, simpla elaborare a datelor statistice constituie o muncă zadarnică, o moară care se învîrtește în gol. Acesta este riscul empiriștilor puri" *. Noțiunile de statistică și de probabilități formează aceeași linie de gîndire, care oferă cadrul conceptual capabil să depisteze structurile ascunse sub mulțimea faptelor observate. În sensul modern al termenului, statistica este privită, tot mai mult, ca „știința interpretării faptelor prin raportare la modele probabiliste".

Diferitele etape ale demersului statistic sînt interdependente. Observarea condiționează elaborările ulterioare: date primare inexacte sau fragmentare vor face iluzorii eforturile prelucrării și analizei. Cînd observațiile de bază „nu sînt făcute cu deplină corectitudine, toată elaborarea ulterioară, oricît de savantă ar fi, oricît am recurge, în prelucrare, la tehnici extrem de pretențioase, ca de pildă cea a organizării unei «bănci de date» și o matematizare a lor cu mijloacele mașinilor electronice de calcul, totul nu ar fi decît o pierdere de vreme și o autoinducere în eroare... Și, din păcate, cifrele statistice pe care ne bazăm sînt deseori regretabil de neserioase; uneori chiar inventate cu o lipsă totală de răspundere, dacă nu chiar cu rea-credință" **. Metodele de inferență statistică nu sînt aplicabile decît în condiții particulare, cîteodată foarte restrictive; ele nu sînt pertinente decît dacă se sprijină pe observații îndeajuns de numeroase, bine grupate și bine descrise. Invers, o elaborare chiar corectă a datelor statistice nu poate da rezultate, atunci cînd integrarea în modele probabiliste este inadecvată.

* O. ONICESCU, *Carrado, Gini—statistician umanist,* în *Sociologia militans* (II), Editura științifică, București, 1969, p. 111.

** HENRI H. STAHL, *Teoria și practica investigațiilor sociale,* vol. I, Metode și tehnici, Editura științifică, București, 1974, p. 186.

Scanned with OKEN Scanner

Demersul statistic trebuie condus în spiritul științei din al cărui domeniu material face parte fenomenul investigat. Eficiența sa postulează, ca un principiu fundamental, păstrarea prezentă în gîndire a naturii fenomenului și, numai în al doilea rînd, considerarea proprietăților formale ale procedeelor de aplicat. Conjugarea analizei cantitative și analizei calitative evită, în tratarea matematică a datelor statistice, exagerările pur formale, care au fost și mai sînt și astăzi practicate. Ca o declarație polemică împotriva unor atare exagerări este interpretată butada lui C. Gini: „Statistică cu cît mai puțină matematică în loc de statistică cu matematică". [59]

Statistica este un puternic instrument de cunoaștere; metoda sa se numără printre metodele particulare cu un grad mare de generalitate. În prezent, se cuvine recunoscut că disciplinele, care se vor științifice, devin aproape toate tributare statisticii. [60] Este însă important să se înțeleagă cît de primejdioase pot fi procedeele statistice mînuite cu ușurință. „Statistica — s-a spus cu drept cuvînt — este una din acele științe ai căror slujitori trebuie să dea dovadă de o adevărată artă a discernămîntului" [134, p. 23]. [61]

Demersul statistic presupune, ca orice demers științific, o teorie; el nu poate fi întreprins decît pornind de la această teorie și nu poate fi înțeles decît de cel care cunoaște teoria. Este un fapt epistemologic foarte important, care se regăsește în toate științele și se impune a-l sublinia mereu în una din științele cele mai concrete: statistica.

1.4 STATISTICA ÎN CONDIȚIILE REVOLUȚIEI ȘTIINȚIFICE ȘI TEHNICE CONTEMPORANE

1.4.1 Statistica și noile discipline științifice

În condițiile creșterii exponențiale a științei — una din trăsăturile cele mai pregnante ale lumii contemporane — și ale sporirii caracterului inter- și multidisciplinar al acesteia, capătă o importanță aparte determinarea locului statisticii în contextul actual, definirea specificității obiectului și metodelor sale. După estimațiile experților UNESCO, numărul disciplinelor științifice s-ar ridica, în prezent, la 1 200; nu este nici o îndoială că numărul lor crește sistematic și rapid. Realitatea lumii, societatea și economia devin tot mai

[59] Butada lui Gini mergea alături cu cea a lui B. Russell: „Matematica este știința în care nu se știe niciodată despre ce se vorbește și nici dacă ce se spune este adevărat sau fals", pe care Gini obișnuia să o citeze în focul polemicilor sale. Vezi O. Onicescu, *op. cit.*, p. 116.

[60] Așa cum Ronald Fisher a remarcat: „Către statisticieni ne îndreptăm pentru a răspunde la problemele cele mai serioase puse de marile aventuri ale epocii noastre. De fiecare dată cînd se întreprinde ceva important, statisticienii sînt în culise. În toate cercetările științifice de bază, ei sînt cei care elaborează planurile de experiență sau observații și tot la ei se recurge pentru a analiza rezultatele, pentru a evalua constatările și pentru a separa faptele clar demonstrate de cele ce mai cer confirmare".

Citat după: *Les sciences sociales dans l'enseignement supérieur. La statistique*, UNESCO, Paris, 1957, p. 14.

[61] Cît de instructive sînt în această privință *Povățuirile...* lui Ion Ionescu de la Brad: „Însă aceea ce mai înainte de toate este necesar statisticianului, în adevăr vrednic de acest nume, este facultatea observației, arta și mai ales tactul de a deosebi aceea ce este adevărat de aceea ce este mincinos în cifrele ce i se înfățișează, probitatea științifică și neînvinsă ce-l face să caute adevărul și numai adevărul... Din lipsa acestei calități mulți sînt făcători de statistică, însă foarte puțini de acei ce inspiră confienă prin inteligența, nepărtinirea și caracterul lor" [59].

Scanned with OKEN Scanner

complexe şi solicită abordări din diverse unghiuri de vedere: unul şi acelaşi fenomen, unul şi acelaşi domeniu al vieţii devine obiect de studiu al mai multor ştiinţe, un „condominium" al mai multor discipline. Diviziunea ştiinţei presupune însă în permanenţă folosirea unor concepte şi metode comune, după cum diversitatea realului cere sistematic analiză şi sinteză.

Este caracteristic pentru timpurile noastre faptul că se constituie o ştiinţă de maximă generalitate, teoria sistemelor, bazată pe însuşirile de sistem pe care le întîlnim în cele mai variate domenii şi, pe de altă parte, statistica — mai precis, metoda şi conceptele sale de bază — se aplică pe scară tot mai largă, se desfăşoară cu multă rapiditate procesul de „statisticizare" a ştiinţelor. Dacă ne referim mai ales la teoria sistemelor, la teoria informaţiei şi la cibernetică, vom remarca prezenţa autoritară a statisticii în cadrul acestora. Desigur este vorba de o statistică ea însăşi energic modernizată. În general, tendinţa sistemică şi cea de cuantificare se afirmă tot mai mult în cîmpul ştiinţei moderne. Se apreciază că din inventarul ideilor de bază, intervenite în ştiinţele sociale între anii 1900 şi 1965, inventar care numără 62 de metode, concepte sau şcoli fundamentale, numai nouă au rezultate necantitative, toate celelalte conducînd la probleme explicit sau implicit cantitative.

Ajutorul solicitat statisticii de către ştiinţă este facilitat prin aptitudinea ei de acomodare la caracterul procesului investigat, prin eficacitatea şi economicitatea procedeelor sale. Viaţa şi politica economică şi socială a zilelor noastre nu mai pot fi concepute fără ea. „Statistica afişează viaţa noastră în mii de modalităţi" [80]. Cum se explică omnivalenţa acestui instrument atît de solicitat, în continuă perfecţionare?

Sub ochii noştri, procesul de înnoire a teoriei şi de reevaluare a metodologiei statistice continuă, ca urmare: a) a creşterii aportului pe care statistica trebuie să-l aducă la dezvoltarea societăţii contemporane; b) a apariţiei unor noi discipline, izvorîte unele pe însuşi terenul statisticii, altele folosindu-se din plin de statistică, dar mai ales a valorificării unor ramuri noi ale matematicii şi în special a modernizării statisticii matematice; c) a impactului informaticii, care nu numai a revoluţionat metodele de prelucrare a informaţiei, dar a condus şi la o nouă concepţie asupra acesteia şi asupra rolului său în conducerea social-economică [22].

Una din caracteristicile vremii noastre o constituie procesul de osmoză şi interferenţă dintre ştiinţele naturii şi ale societăţii, ceea ce a condus la apariţia şi dezvoltarea aşa-numitelor discipline de intersecţie sau de frontieră. Pentru cunoaşterea şi analiza fenomenelor din natură şi societate avem astăzi la dispoziţie diverse discipline şi teorii, care se situează aproape de/şi în vecinătatea statisticii.

Considerarea ştiinţei ca forţă nemijlocită de producţie a căpătat în zilele noastre un caracter axiomatic; aceasta înseamnă, în primul rînd, schimbarea funcţiei sociale a ştiinţei, materializarea cunoştinţelor ştiinţifice în elemente materiale ale forţelor de producţie şi în formele de organizare corespunzătoare; ştiinţa devine baza teoretică a proceselor de producţie. „Ştiinţa devine o forţă nemijlocită de producţie pe trei căi: a) prin întruchiparea ei *materială*; b) prin întruchiparea ei în *forţă de muncă*, în factorul subiectiv al producţiei, prin ridicarea pregătirii ştiinţifice şi perfecţionarea spirituală a omului; c) prin *conducerea* proceselor sociale tot mai complexe şi prin *organizarea muncii* şi a producţiei" [112].

Scanned with OKEN Scanner

În zilele noastre, sub impactul condițiilor complexe generate de procese de natură foarte diversă și supuse transformărilor rapide, intervenția științei în procesul de conducere constituie o necesitate. A devenit astfel posibilă și necesară *știința conducerii*, aflată în plin proces de dezvoltare ca „un ansamblu de postulate, de principii (teoria conducerii), de elemente de orientare în studiul fenomenelor de conducere (metodologia conducerii) și de mijloace de acționare în vederea atingerii unui țel (tehnica conducerii)... Știința conducerii are menirea să facă efortul de a integra cunoștințe din domenii diferite și de a promova o viziune interdisciplinară". [62]

Apariția *ciberneticii*, legată de numele lui Norbert Wiener, a marcat un moment de importanță deosebită în procesul de fuziune dintre acțiunea socială și conducere, de transformare a conducerii în știință. Constituindu-se ca un ansamblu de principii ale comunicației, controlului și comenzii și concretizîndu-se ca teorie a sistemelor cu autoreglare, cibernetica oferă aparatul său conceptual pentru edificarea metodologiei științei conducerii, plecînd de la noțiunea fundamentală de sistem și tratînd organismele social-economice ca *sisteme cibernetice*.

Activitatea de conducere constituie un proces complex de cuncaștere, judecată și acțiune și ca atare ea se concretizează în operații de informare și analiză (unde intervenția statisticii este indispensabilă), decizie și control, punctul de greutate constituindu-l momentul luării deciziei. Astfel, s-a ajuns la constituirea unei *teorii a deciziei* pe bază statistico-matematică, pentru fundamentarea științifică a procesului decizional.

Teoria deciziei se integrează în ansamblul de teorii, metode și tehnici, cunoscut sub denumirea de cercetare operațională. Considerată drept instrument de conducere rațională a unei activități complexe, cercetarea operațională integrează, pe lîngă teorii și procedee matematice și statistice, un întreg complex de măsuri în care se îmbină metode științifice de organizare și conducere, mijloace tehnice adecvate, metode perfecționate de culegere și prelucrare a informației.

Perfecționarea *tehnicii de calcul* își pune și ea o amprentă cu consecințe de netăgăduită importanță asupra evoluției statisticii.

Calculul manual, folosit în prelucrarea datelor încă de funcționarii faraonilor egipteni și mecanizat odată cu introducerea, în anii '80 ai secolului trecut, a mașinii electrice de perforat Hollerith, dispare astăzi complet prin perfecționarea mecanizării și automatizării lui. După afirmația lui Henri Guitton, analiza statistico-economică de dinaintea apariției calculatoarelor nu mai este de aceeași natură cu cea de după apariția lor, deoarece calculatoarele au creat posibilitatea să se abordeze noi tipuri de analiză, care nu erau de conceput cu tehnica de calcul anterioară. Aceasta permite efectuarea de calcule pentru cele mai complicate sisteme de ecuații, iar prelucrarea electronică va determina și o sporire a exigențelor față de practică, atît în ce privește complexitatea și varietatea datelor, cît și în privința tezaurizării și regăsirii lor [22].

[62] S. TAMAȘ, *Știința conducerii societății în socialism*, Editura politică, București, 1974, p. 15 și 24.

Scanned with OKEN Scanner

Perfecționarea tehnică a prelucrării datelor statistice se desfășoară concomitent cu dezvoltarea unor procedee mai eficiente de culegere și înregistrare a datelor, mai ales prin perfecționarea procedeelor de eșantionare.

1.4.2 Progrese recente în statistică

În ultimele decenii, statistica matematică a înregistrat mari progrese care au înrîurit puternic statistica.

Teoria proceselor stohastice a devenit, în prima parte a secolului nostru, un puternic instrument de cercetare a fenomenelor din natură și societate. N. Arhey a aplicat-o la particulele întîlnite în radiațiile cosmice, G. Malecot în experimentele genetice, iar Ragnar Frisch la studiul ciclurilor din economie[63].

Axiomatizarea teoriei probabilităților de către A. N. Kolmogorov, în 1933 [85], a dovedit că probabilitatea nu reprezintă doar o valoare teoretică ce decurge din calcule formale, ci ea rezultă din studiul adîncit al fenomenelor și proceselor din natură și societate, din legile obiective de desfășurare a fenomenelor de masă, fiind o consecință a lor, adeverind astfel consistența principiului științific al legăturii dialectice dintre teorie și practică. Academicienii Octav Onicescu și Gheorghe Mihoc au lărgit cadrul axiomatic al teoriei probabilităților prin așa-numitele *funcții sumă*, care răspund necesității de reprezentare a mărimilor corespunzătoare evenimentelor cîmpului și prin care se pot enunța și demonstra principalele teoreme ale teoriei probabilităților.

Unii probabiliști români, precum I. Cuculescu și Gh. Marinescu, au extins utilizarea funcțiilor sumă.

Teoria testului secvențial, ca și teoria deciziei, — a cărei apariție și dezvoltare au mers în pas cu evoluția calculului probabilităților și a statisticii matematice și care poate fi definită drept „știința complexă a hazardului" — au dat impulsuri noi statisticii. Ambele au fost inițiate de Abraham Wald.

În matematică, teoria deciziei, numită teoria funcțiilor de decizie, a început să se dezvolte prin așa-numitul principiu minimax, introdus în 1895 de H. Minkovski. În anul 1921, Emile Borel a găsit posibilitatea aplicării acestui principiu la aspectele stohastice ale teoriei jocurilor și la condițiile din economie. Această utilitate „strategică" a teoriei minimax a fost dezvoltată în anii următori de John von Neumann, care a valorificat-o în analiza statistico-matematică a unor fenomene economice (producție și prețuri), iar în 1944, odată cu apariția lucrării *The Theorie of Games and Economic Behaviour* (Teoria jocurilor și comportamentul economic), de John von Neumann și Oscar Morgenstern, începe să se definească o nouă disciplină de intersecție. „Teoria statistică a decizei — după G. Menges — se situează ca o disciplină de cercetare la marea intersecție dintre statistică, matematică și științele economice" [80, v. 1]. A. Wald a interpretat problema decizională ca un joc cu sumă nulă între doi parteneri, în sensul teoriei jocurilor, dezvoltînd de aici teoria funcțiilor statistice de decizie. După aceea a fost încadrată în teoria statistică a deciziei, formulată de Wald, întreaga teorie statistică clasică, aplicată la testarea ipotezelor și la estimarea limitelor de încredere.

Aplicarea teoriei deciziei presupune formularea unor criterii după care se examinează problema dată. Luarea în considerare, simultan, a mai multor criterii de decizie îi conferă caracterul de multidimensionalitate. De asemenea s-a conturat necesitatea unei analize multidimensionale, numită și analiză

Scanned with OKEN Scanner

multicriteriu, prin care se urmăreşte degajarea, din multitudinea de date obţinute într-o perioadă de timp şi organizate în serii statistice, a „axelor factoriale" ale fenomenelor respective, cu structura lor majoră şi minoră. Spre deosebire de alte tipuri de analiză statistică, de exemplu analiza factorială, unde punctul de plecare îl constituie anumiţi factori stabiliţi aprioric, aici aceştia sînt căutaţi, iar unităţile observate sînt stări succesive ale unui fenomen economic, sesizat de-a lungul unor multiple dimensiuni [22; 51].

1.4.3 Statistica şi modelarea

Cunoaşterea proceselor social-economice rezultă din observarea metodică comparativă, în cursul căreia ele pot fi comensurate (cuantificate). Puterea de abstracţie se reflectă în *modelarea* fenomenelor şi proceselor, care se îmbină cu desfăşurarea analizei cantitative, ca bază pentru efectuarea analizei calitative. „Elaborarea modelelor nu elimină statistica din cîmpul cercetării ştiinţifice şi nici nu-i substituie alte discipline, ci, din contră, o solicită în mod intensiv pe mai multe planuri" [22]. Modelul ajută statistica să pătrundă, după expresia lui M. G. Kendall, în „camera obscură" *(black box)*, ceea ce înseamnă a privi în mod efectiv şi a căuta identificarea interrelaţiilor şi nu este vorba de relaţii doar promiţătoare pline de elemente stohastice şi de funcţii necunoscute.

Metoda modelării a dobîndit o importanţă considerabilă odată cu intensificarea aplicării metodelor statistico-matematice. „Primul asistent al oricărui conducător este statisticianul sau oficiul de culegere şi prelucrare a datelor. Nu există legătură între sistem şi conducere dacă nu există între ei un nucleu organizat pentru culegerea de date... Dar strîngerea de date nu epuizează operaţia conducerii. Datele trebuie organizate. Ele sînt consolidate într-un model. Modelul este o machetă a realităţii şi este pe masa conducerii" [63].

Modelarea constituie o modalitate specifică de informare sau cunoaştere, un procedeu şi o formă de organizare a cunoaşterii. Modelarea constituie o cale de efectuare a experimentelor în domenii în care experimentul este limitat sau exclus. În metodologia previzională, modelele econometrice de tip statistic întocmite astăzi şi care urmăresc relevarea structurii datelor statistice şi nu simpla corelare a lor prezintă un progres faţă de metodele anterioare şi asigură rezultate predictive mai bune. Elaborarea de modele ale economiei naţionale a contribuit în mod direct la perfecţionarea sistemului statistic în numeroase ţări; este cazul balanţelor legăturilor dintre ramuri.

1.4.4 Statistica şi informatica

Informaţia are în statistică un rol central, ceea ce l-a făcut pe Günther Menges să considere statistica drept „conceptul reprezentativ al informaţiei" [80, 1].

Organizarea obţinerii, transmiterii şi utilizării informaţiei constituie un proces în ale cărui etape de desfăşurare se identifică fenomenele, respectiv evenimentele care generează informaţia, se stabileşte purtătorul material al informaţiei şi modul în care se culege şi se înregistrează informaţia, se alege

[63] M. MALIŢA, *Aurul cenuşiu*, vol. I, Editura Dacia, Cluj-Napoca, 1971, p. 92—93.

Scanned with OKEN Scanner

metoda şi instrumentul de prelucrare şi păstrare a ei, se defineşte destinaţia şi se organizează transmiterea informaţiei la acea destinaţie. Indiferent dacă procedeul de efectuare a acestor operaţii este manual, mecanic sau electronic, elementele de mai sus nu pot lipsi din circuitul informaţional organizat. Astfel a luat naştere *sistemul informaţional*, devenind, în cadrul organizatoric statal, instrument al conducerii. Perfecţionarea formelor de conducere şi organizare economico-socială, în consens cu mecanizarea şi automatizarea proceselor tehnologice, impune modernizarea la un nivel corespunzător a procedeelor de tratare a informaţiei. Automatizarea ei facilitează, adînceşte şi precizează considerabil cunoaşterea realităţii şi deschide calea asimilării sistemului informaţional unui flux tehnologic, prin transformarea lui într-un sistem cibernetic; se spune că sistemul informaţional devine sistem *informatic.*

Perfecţionarea conducerii în conformitate cu stadiul actual al dezvoltării ştiinţei şi tehnicii înseamnă introducerea tot mai accentuată în mecanismul ei a principiilor şi tehnicii *informaticii*, pentru asigurarea unui suport raţional în elaborarea deciziilor şi a operativităţii în dirijarea activităţii. Informatica se interferează cu cibernetica, teoria deciziei şi alte discipline, fundamentînd prin asocierea unor principii pluridisciplinare construirea bazei teoretice şi metodologice a ştiinţei conducerii. Amplificarea funcţiei informaticii a lărgit sfera ei conceptuală iniţială cu caracter eminamente tehnic. Informatica este privită în prezent ca disciplina ştiinţifică a prelucrării automate a informaţiei[64], dezvoltînd ideea de informaţie pe baza unor principii cibernetice, în cadrul teoriei sistemelor.

Introducerea automatizării în prelucrarea datelor constituie un salt calitativ important, o adevărată revoluţie, cu consecinţe ce încă nu se pot bănui în întregime şi asupra statisticii. Calculatorul electronic conduce la o nouă viziune cu privire la funcţiile sociale ale informaticii, implicit şi cu privire la statistică, precum şi la căile de realizare a acestor funcţii. „Dacă în concepţia veche, clasică, accentul se punea pe consemnarea evenimentelor într-o viziune statică paseistă, a cărei expresie sintetică o constituia conceptul de evidenţă a economiei naţionale, cu cele trei forme..., în concepţia modernă, bazată pe conceptul de sistem informaţional, accentul se pune pe prelucrarea şi transmiterea datelor, pe organizarea raţională a fluxului informaţional — care să asigure o informare completă şi rapidă a factorilor de decizie, într-o viziune prospectivă“ [22].

Automatizarea prelucrării datelor statistice nu afectează funcţia ştiinţei şi metodologia statistică. Ele nu se pot substitui pentru că obiectul, metodele şi funcţiile lor, caracteristicile lor specifice sînt diferite. „Combinarea metodelor statistice cu cele ale informaticii este de natură să îmbogăţească simţitor atît tehnicile de prelucrare a datelor, cît şi calitatea şi conţinutul informaţional al indicatorilor statistici... Sub incidenţa informaticii, statistica este obligată să evolueze în direcţii noi, să-şi modifice într-o anumită măsură modul său tradiţional de acţiune“ [22].

Este semnificativă, din acest punct de vedere, evoluţia statisticii către ceea ce acad. Octav Onicescu numeşte „statistică informaţională“ şi pentru

[64] După definiţia din 1967 a Academiei franceze, informatica este „ştiinţa prelucrării raţionale, îndeosebi prin maşini automate, a informaţiei, considerată ca suport al cunoştinţelor umane şi ale comunicărilor în domeniile tehnice, economice şi sociale“. După *Mic dicţionar enciclopedic* (Ed. enciclopedică română, 1972): „ştiinţa care se ocupă cu studiul prelucrării informaţiei cu ajutorul sistemelor automate de calcul“.

Scanned with OKEN Scanner

care d-sa a şi elaborat indicatorii „energie informaţională" şi „corelaţie informaţională", care şi-au găsit foarte repede domenii diferite de aplicare atît în ţara noastră, cît şi în străinătate.

Calitatea de instrument statistic a energiei informaţionale a apărut şi la calcularea coeficientului de corelaţie informaţională, deschizînd — aşa cum arată acad. Octav Onicescu — un drum statisticii matematice şi aplicaţiilor ei în variate domenii, mai ales în cele pur calitative, în care informaţia este dată de frecvenţe, şi nu de valoarea unei mărimi.

Teoria matematică a informaţiei, care a intrat în arsenalul metodologic al ciberneticii şi informaticii, formulează legile generale ale comenzii, controlului şi comunicaţiilor şi stabileşte principiile de codificare, prelucrare, păstrare şi transmitere a informaţiei şi se asociază cu tehnica de tratare automată. Ca moment de cotitură în dezvoltarea acestei teorii se consideră apariţia, în anul 1948, a lucrării lui Claude E. Shannon, *O teorie matematică a transmiterii informaţiilor* [65], în care autorul se ocupă de cuantificarea informaţiei transmise şi elaborează, pe baza teoriei şi a calculului probabilităţilor, metoda matematică de reducere a perturbaţiei, care influenţează transmiterea informaţiei. Cu zece ani mai tîrziu, A. K. Kolmogorov a scos în evidenţă locul central pe care conceptul de informaţie îl ocupă în cibernetică, definind-o totodată pe aceasta mai cuprinzător, adică drept „ştiinţa modurilor de recepţie, păstrare, transformare şi folosire a informaţiei în maşini, organisme vii şi reuniuni ale lor" [66]. Prin identificarea relaţiilor de comunicaţie între celulele, respectiv organele dintr-un organism (fie el biologic, mecanic sau social) se asigură funcţionarea şi stabilitatea sa faţă de influenţele (perturbaţiile) din afară.

Prezintă interes, fără îndoială, examinarea raportului dintre statistică şi alte discipline moderne. În peisajul actual al ştiinţei, raportul cu teoria sistemelor, cibernetica şi informatica are importanţă majoră. De fapt, statistica este prezentă cu metode noi şi în discipline ca: sociometria, econometria, psihometria, demografia, marketingul, cărora le asigură un progres considerabil aplicativ.

Ca o concluzie preliminară se poate menţiona că, în cadrul relaţiilor de colaborare dintre ştiinţe, beneficiul este comun: al statisticii, care îşi îmbogăţeşte substanţial conceptualizarea şi metodologia, al disciplinelor colaboratoare, care îşi lărgesc şi adîncesc cadrul de aplicaţii. Înţelegerea acestui proces reclamă însă o cunoaştere amănunţită a metodelor statisticii, ceea ce constituie obiectul capitolelor următoare.

[65] CL. E. SHANNON, *A mathematical theory of communication*, în „Bell System Technical Journal", 1948.

[66] A. N. KOLMOGOROV, *Kibernetica*, în *Bolşaia Sovietskaia Enţiclopedia*, vol. 51, Moscova, 1958.

Scanned with OKEN Scanner

CAPITOLUL 2
Procesul cunoașterii statistice

2.1 INFORMAȚIA ȘI SISTEMUL INFORMAȚIONAL STATISTIC

Caracterul specific al obiectului și metodei statisticii își pune amprenta asupra demersului statistic, asupra fiecăreia dintre treptele prin care se realizează această modalitate caracteristică de cunoaștere, de la descriere la analiză. Ceea ce este însă semnificativ, este că modernizarea cuprinde toate fazele, fără excepție. Desigur, și în viitor, se va vorbi de observarea statistică, de prelucrarea infor mațieistatistice, de analiză, dar într-o conceptualizare absolut nouă. Diferitelor tipuri de evidență, așa cum se cunoșteau cu 15—20 de ani în urmă, le-a luat locul sistemul informațional și, mai recent, sistemul informatic. Locul recensămîntului și altor cercetări statistice clasice îl iau acum băncile de date, registrele permanente de populație. Prelucrarea manuală, pe care generații de statisticieni au practicat-o, a fost înlocuită cu prelucrarea mecanizată, iar în ultimii ani, cu prelucrarea automatizată sau electronică. Totul se află sub semnul aceleiași tendințe inovatoare. Marile realizări și progrese ale teoriei informației, ciberneticii, teoriei sistemelor și ale altor discipline sînt preluate și încorporate statisticii. De aceea, în cele ce urmează, unei expuneri „clasice" a fazelor cunoașterii statistice și a modalităților sale îi este preferată o inventariere sumară a cîtorva momente mai semnificative, dar care ține seama de modernizarea statisticii, evocîndu-se succint acele elemente care țin astăzi mai curînd de tradiția ei. Prin urmare, este justificată prezentarea mai amplă a sistemelor informatice sau a prelucrării electronice în dauna unor procedee cum sînt rencensămintele sau monografiile statistice. Dealtfel, însă și reprezentarea clasică este puternic amendată: informația este captată, transmisă sub formă prelucrată unui beneficiar (decident) care, prin deciziile sale, generează noi fluxuri informaționale.

Întrucît funcția statisticii este cercetarea fenomenelor și proceselor din natură și societate și punerea în evidență a legităților lor de manifestare, cunoașterea acestora se dobîndește prin observarea de masă și se exprimă în date sau informații cuantificate.

Pentru informația statistică se folosea pînă de curînd, în mod obișnuit, termenul de dată sau date, dar intrarea foarte rapidă în circulație a noțiunilor folosite în cibernetică și accepțiunea foarte complexă dobîndită în ultimul timp de noțiunea de informație fac necesară distincția între informație și dată, în sensul că termenul din urmă este legat mai mult de operația de pre-

Scanned with OKEN Scanner

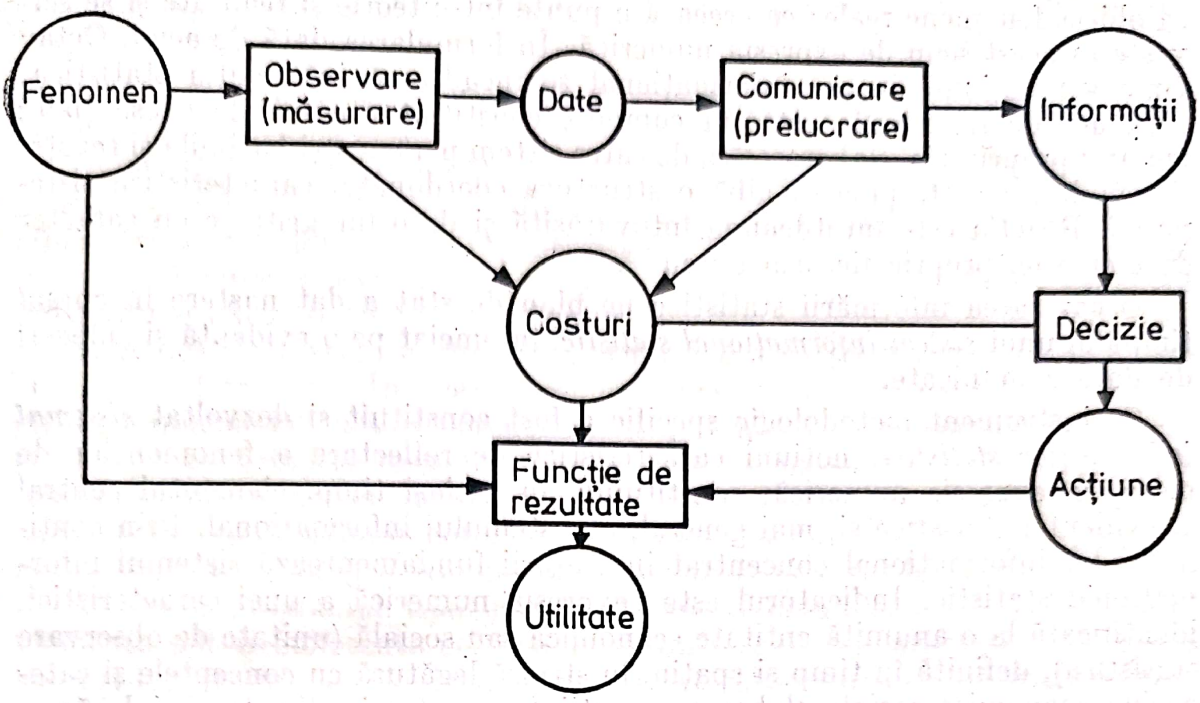


Fig. 2.1 Schema de bază a sistemului informațional.

lucrare. De pildă, cibernetician sovietic I. I. Cerneak [1] precizează că „datele constituie materia primă din care se extrag informațiile, ... ele există obiectiv ..., fiind fixate pe anumiți purtători și circulă în interiorul sistemului, în procesul de conducere".

G. Menges — H. J. Skala [80. v. 2] arată: „... noi tratăm obținerea și prelucrarea datelor ca părți ale unui mare *sistem informațional* economic ... pentru că statistica este un concept reprezentativ de date". În cadrul sistemului informațional datele reprezintă simboluri de ieșire (output) ale unei prime transformări sau, în sensul teoriei informației, ale unui „canal" adică al *observării.* Acest canal are ca simboluri de intrare (input) *realizări* sau *comensurări* ale unui fenomen. Inputul este transformat prin înregistrare (în fazele nemijlocit antecedente sau următoare ale planului observării) și prezentate în date. Printr-un sistem de transformări ulterioare, datele sînt prelucrate, adică sînt comprimate și interpretate în sensul „realității și transmise unui *decident,* care le primește sub formă de *informație* și le folosește pentru decizii, ceea ce înseamnă că prin decizia sa ele sînt transformate în *acțiuni.* Acestea pot fi anumite măsuri cu caracter politic ori luate în cursul activității din cadrul întreprinderilor ... sau constituie decizii statistice, în sensul unor testări, a unor ipoteze sau prognoze".

Schematic, sistemul informațional astfel conceput poate fi reprezentat ca în fig. 2.1.

Desigur, informația statistică nu acoperă decît în parte conceptul de informație [2]; ea trebuie să fie orientată după modelul teoretic, să fie teoretic fundamentată. Prin urmare, informația statistică trebuie să aibă

[1] I. I. CERNEAK, *Elaborarea sistemului informațional economic,* trad. din limba polonă, „Buletin Documentar Statistic" nr. 5/1969.

[2] Se pare că termenul de informație a fost utilizat pentru prima dată de Georg v. Mayr, pentru definirea conceptului de statistică, într-o lucrare a sa din 1914.

Scanned with OKEN Scanner

ca obiect fenomene reale; ea creează o punte între teorie şi realitate şi se serveşte în acest scop de expresia numerică. În formularea dată de acad. Octav Onicescu noţiunii de act informaţional se încadrează informaţia statistică: „Un act de informaţie este un complex constituit din trei faze esenţiale: mesajul propriu-zis, elaborarea sa de către sistem pentru a şi-l asimila şi reacţia sistemului, care trebuie să aibă o structură coordonată, caracteristică sistemului. Reacţia este întotdeauna întovărăşită şi de o înregistrare cu caracter de memorie, proprie fiecărui sistem“[3].

Organizarea informării statistice pe plan de stat a dat naştere în cursul timpului unui *sistem informaţional statistic*, întemeiat pe o evidenţă şi culegeri de date comunicate.

Ca instrument metodologic specific a fost constituit şi dezvoltat *sistemul indicatorilor statistici*, noţiuni caracteristice de reflectare a fenomenelor de masă, în expresie numerică, constituind, în acelaşi timp, elementul central al evidenţei statistice şi, mai general, al sistemului informaţional. Prin conţinutul lor informaţional concentrat indicatorii fundamentează sistemul informaţional statistic. Indicatorul este „expresia numerică a unei caracteristici, identificată la o anumită entitate economică sau socială (unitate de observare statistică), definită în timp şi spaţiu, în strînsă legătură cu conceptele şi categoriile economico-sociale elaborate de ştiinţa sau teoria din domeniul căreia face parte entitatea respectivă“ [23].

În concepţia modernă, sistemului informaţional statistic, constituit din semnale caracterizate pe trei trepte principale de desfăşurare — observarea, transmiterea şi decizia —, i se atribuie următoarele caracteristici fundamentale: manifestările măsurate ale fenomenelor sînt transformate în date prin înregistrare, datele devin informaţii prin transmitere, respectiv prelucrare, informaţiile sînt transpuse în acţiuni prin decizii. Utilitatea acţiunii testează valoarea sistemului informaţional statistic [80, v. 2].

Sistemul informaţional statistic deţine o poziţie cheie în sistemul informaţional de conducere, sfera informaţiei statistice fiind cu mult mai largă decît cea a celorlalte categorii de informaţii necesare conducerii. Ea însoţeşte procesul de conducere, în toate etapele sale: planificare, execuţie, control.

Sistemul informaţional statistic în accepţia sa clasică, bazat pe evidenţa statistică şi pe sistemul dărilor de seamă periodice, a jucat timp îndelungat un rol important; astăzi însă el nu mai corespunde pe deplin necesităţilor conducerii.

Introducerea automatizării prelucrării datelor şi crearea de sisteme informatice integrate de conducere a activităţii implică renunţarea, în perspectivă, la sistemul tripartit al evidenţei economice — tehnic-operativă, contabilă şi statistică — şi introducerea unui sistem cu totul nou de culegere, prelucrare, păstrare şi valorificare a informaţiei. A şi luat fiinţă, în unele ţări socialiste, aşa-numitul sistem unitar de calcul economic şi statistică, ale cărui obiective depăşesc sfera simplei observări şi înregistrări cifrice a manifestării fenomenelor şi proceselor reale.

[3] O. ONICESCU, *Informaţie şi sistem în legătură cu raportul Meadows*, „Revista de Statistică“ nr. 2/1973.

Scanned with OKEN Scanner

2.2 OBSERVAREA STATISTICĂ

Desfășurarea unei cercetări statistice are — așa cum s-a mai spus — ca primă treaptă *observarea de masă.*

Datele obținute în urma observării fiecărei unități a colectivității cercetate constituie materialul statistic inițial, „materia primă", pe baza căreia se obține cunoașterea statistică. Numai o înregistrare statistică, în care realitatea este investigată corect și exact, poate asigura o analiză de bună calitate, concluzii autentice și eficiente [88].

Orice fel de cercetare statistică sistematică necesită organizarea desfășurării ei planificate. În acest scop urmează să fie rezolvate, în prealabil, cîteva probleme cu caracter predominant de tehnică a desfășurării observării, și anume: a) stabilirea metodei de observare din punctul de vedere al extinderii ei asupra unităților de observare, b) determinarea procedeelor observării și c) efectuarea de experimente, ca metodă de observare. Planul de organizare și efectuare a observării trebuie să precizeze scopul și obiectul ei, unitatea de observare, programul, tipul și locul observării, detalii organizatorice. Aceste elemente pot fi cunoscute cel mai bine din activitatea practică și de aceea nu insistăm asupra lor. Amintim însă că există observarea totală și observarea parțială sau prin sondaj.

Pentru cunoașterea fenomenelor social-economice, în care predomină statistica descriptivă, observarea totală (de masă) constituie metoda caracteristică. Ea are însă dezavantajul că este costisitoare, în ce privește cheltuielile materiale și necesitățile de personal calificat, nu poate fi efectuată oricînd și oriunde și reclamă timp, ceea ce afectează de multe ori actualitatea și utilitatea cunoașterii. Din acest motiv, observarea totală, instrument clasic de cunoaștere statistică, face tot mai mult loc în zilele noastre observării parțiale, care, putînd acorda mai multă atenție unui număr mai mic de unități, conduce adesea la rezultate mai corecte.

Observarea parțială, mai ales în varianta sa de sondaj, se dezvoltă pe scară tot mai largă în prezent, nu numai datorită accesibilității sale ca instrument de cunoaștere în cele mai diferite domenii, dar și mulțumită faptului că, aplicată parțial, creează posibilitatea utilizării mai aprofundate a metodelor matematice noi și a tehnicii automate de calcul. Ca atare, există în prezent procedee bine puse la punct de desfășurare a observării prin sondaj, asupra cărora se va stărui pe larg în capitolele următoare.

Experimentul statistic. Definit ca un procedeu de cercetare în știința modernă și constînd în reproducerea artificială sau modificarea intenționată a unui fenomen natural, cu scopul observării lui în condițiile special create de cercetător, acest instrument de „observare provocată" a depășit în epoca contemporană sfera aplicativă a științelor naturii. El este utilizat tot mai mult și pentru cunoașterea, la scară macro sau micro, a fenomenelor social-economice, pentru studierea legăturilor, a consecințelor acestor procese în viața social-economică.

În teoria și metodologia statisticii, experimentul a fost introdus de către Ronald A. Fisher, sub denumirea de planificare a experimentului *(design of experiments)*, ceea ce ar însemna experimentare statistică programată. Statisticianul intervine chiar de la programarea observării și este prezent

Scanned with OKEN Scanner

nu numai la analiza rezultatului experimentului; instrucțiunile de desfășurare a lui trebuie astfel alcătuite încît să asigure o informare cît mai bogată, cu cheltuieli minime. Asemenea experimente se efectuează adesea pentru a se testa ipotezele statistice, de care ne vom ocupa în unul din capitolele următoare. Experimentul statistic leagă strîns metoda inductivă (inferențială) cu metoda deductivă [80, v. 2] [4].

2.3 PRELUCRAREA MATERIALULUI STATISTIC

Materialul obținut prin observare, în cea de a doua etapă a cercetării statistice, este supus centralizării, grupării și prelucrării datelor.

Gruparea statistică este o metodă de prelucrare, în care populația statistică cercetată este separată în grupe omogene sau clase. În anumite condiții, chiar simpla grupare a datelor statistice poate dezvălui: a) tipurile social-economice și structura colectivității, b) legăturile dintre caracteristici, gruparea prefațînd astfel metoda corelației, c) deplasările sau modificările structurale.

Pentru constituirea claselor și stabilirea intervalelor, se cere mult discernămînt, precum și o analiză calitativă a fenomenelor.

Cînd grupările efectuate dobîndesc un caracter de stabilitate, de durată, se obișnuiește să fie denumite nomenclatoare și clasificări, de exemplu, clasificarea întreprinderilor, ramurilor economiei naționale, personalului etc. În domeniul relațiilor economice cu străinătatea se utilizează nomenclatoare și clasificări standardizate pe plan național și internațional. Ele joacă un rol important în definirea structurii indicatorilor statistici, prin elementele ce le oferă pentru circumscrierea unităților de observare, din punct de vedere teritorial, al organizării în cadrul social-economic (pe profil de ramură, pe produse etc.).

Procedee și instrumente tehnice de prelucrare. Procesul tehnologic de prelucrare a informației statistice cunoaște, în evoluția sa istorică, trei etape: prelucrarea manuală, mecanică și automată. Demarcația între acestea nu este rigidă. Uneori procedee vechi au fost adoptate ca părți componente ale celor noi. Astăzi coexistă toate trei, iar pînă la introducerea echipamentului electronic coexistau primele două, chiar în cadrul aceluiași organism.

Cele mai vechi evidențe scrise, în forma unor scrieri pictografice pe tăbliță de lut apar la premergătorii babilonienilor, vechii sumerieni, în mileniul al treilea înaintea erei noastre. Astfel de tăbliță s-au descoperit mai tîrziu și în Orientul apropiat, ca și în regiunea Mediteranei. La vechii egipteni, evidențele au luat o mare extensiune, fiind ținute de scribi pe foi de papirus, păstrate în suluri; ele cuprindeau informații asupra diverselor operații de vînzare-cumpărare, situații fiscale ș.a. Mai tîrziu au apărut răbojurile, probabil folosite mai mult în Anglia după invazia lui William Cuceritorul, în 1066, deși crestarea sau găurirea bețelor, ca mijloc de înre-

[4] Este locul să remarcăm că statistica matematică, în concepția lui K. Pearson și R. A. Fisher, s-a dezvoltat mai cu seamă pe baza planificării experimentelor, așa cum s-a practicat la Stațiunea experimentală de la Rothamsted, preluată apoi de alte școli naționale (americană, indiană). Cele mai spectaculare aplicații s-au realizat în agrobiologie. În ultimii ani, un domeniu foarte solicitat este cel al controlului calității.

Scanned with OKEN Scanner

gistrare a numerelor, își are originea în neolitic și chiar paleolitic. Perfecționarea evidenței contabile, după introducerea sistemului cu două intrări, la Geneva în 1340, iar în formă tipărită, la Veneția în 1494, în alcătuirea călugărului franciscan Luca Paccioli, a fost urmată mai tîrziu de organizarea evidenței statistice. Inventarea mașinii de scris în secolul al XIX-lea a marcat un important pas înainte în perfecționarea înregistrării și prelucrării datelor.

De o importanță uriașă în domeniul prelucrării a fost perfecționarea procedeelor și dispozitivelor de calcul. Precedînd abacul, pe valea Tigrului și Eufratului, cu vreo 5000 de ani în urmă, a fost imaginată o placă de lut cu canale prin care se deplasau pietricele, ușurînd numărătorile. Acest procedeu a fost apoi perfecționat de chinezi și japonezi. Se crede că abacul, așa cum îl cunoaștem astăzi, a fost inventat în China cam prin 2600 î.e.n. Japonezii foloseau un instrument similar, numit soroban. În Europa, se pare că abacul a început să fie folosit abia de la începutul se colului al 4-lea al erei noastre, mai întîi de către greci, după ce sistemul de scriere cu cifre romane a fost înlocuit prin cel arab, ceea ce a permis să se introducă cifra zero. O contribuție importantă a adus mai tîrziu cartea lui Leonardo Pisano, *Liber abaci*, apărută în 1202. Apoi perfecționarea tehnicii de calcul a stat aproape pe loc pînă în secolul al XVII-lea — „marele secol al progresului matematicii" [120].

Pentru clasificarea datelor se utilizează operații de *codificare*. Acestea reprezintă un sistem de condensare, prescurtare, în scopul reducerii la minimum a muncii necesare înregistrării și transcrierii datelor, dînd în același timp posibilitatea identificării și deosebirii cu mai multă ușurință și rapiditate a datelor. Se folosesc diferite sisteme de codificare: secvențiale, prin atribuirea de numere unei liste de poziții, într-o succesiune riguroasă, începînd cu 1; clasificări codificate pe grupuri de mii, sute, zeci; coduri zecimale, la care numerele sînt atribuite pe o bază zecimală extinsă prin zecimi, sutimi, miimi etc.

Orice informație statistică este înregistrată pe un suport material — purtătorul de informație —, care poate fi un formular, o fișă, o cartelă sau un alt fel de document. În procesul de prelucrare intervine operația de *sortare* a documentelor, adică de grupare a lor după rangul sau după anumite caracteristici; manual sau cu mașini de sortat.

Păstrarea (conservarea, tezaurizarea) datelor în mod sistematic prezintă o deosebită importanță pentru regăsirea și utilizarea lor cu ușurință și rapid. În acest scop se folosesc cutii de anumite dimensiuni, de obicei anumite dulapuri numite *fișiere*, administrarea acestora constînd în operații cu caracter manual, chiar atunci cînd celelalte operații sînt mecanizate sau automatizate. Într-un sistem de administrare a fișierelor, datele sînt clasificate prin *indexare* alfabetică sau numerică, ambele în diverse variante sau combinații între ele, în aranjări mnemonice, care ajută la reținerea lor mintală [8].

Oricît de perfecționată a ajuns în prezent tehnica de prelucrare, operația *manuală*, constituind un proces relativ simplu de desfășurare, continuă a fi folositoare la un volum mic de date ale observării și la anumite verigi de muncă statistică.

Prelucrarea *mecanică* a datelor a fost precedată de perfecționarea tehnicii de calcul, prin inventarea unor dispozitive, care au înlocuit înmulțirea printr-o adunare de logaritmi.

În 1617, matematicianul scoțian John Napier (numit și Neper), inventatorul tabelelor de logaritmi, a alcătuit un aranjament mecanic de bucăți din os, pe care erau imprimate numere; combinîndu-le, Napier a efectuat direct operații de înmulțire. Ca inventator al tabelelor de logaritmi este considerat și matematicianul elvețian J. Bürgi (1552—1632), numit și Bjergius,

Scanned with OKEN Scanner

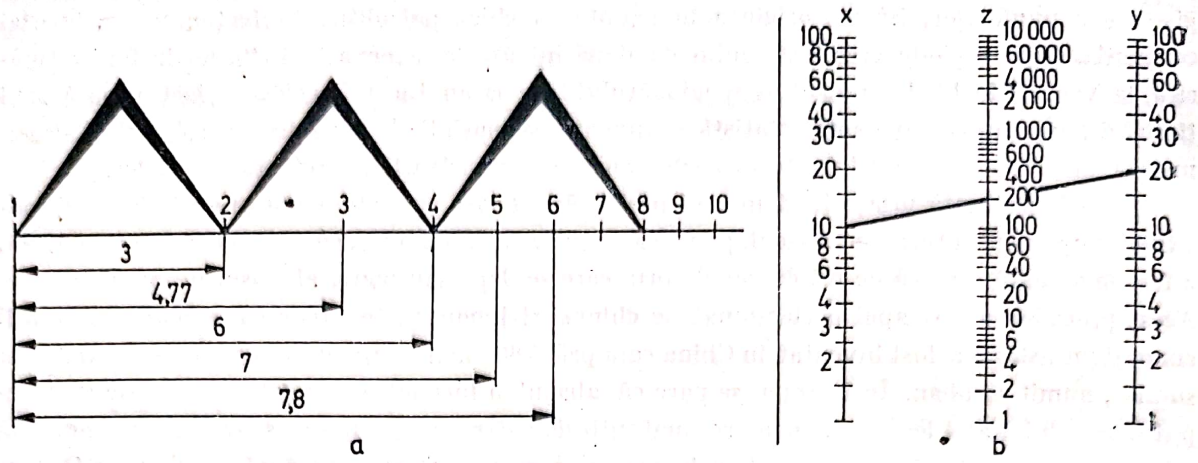


Fig. 2.2 Schema de construire a scării logaritmice (după J. Napier)

care a lucrat separat și independent de Napier. Etimologic, noțiunea de logaritm derivă din termenii greceşti *logos* și *arithmos* (număr), care dau împreună sensul de raport exact. Divizînd logaritmic o dreaptă, J. Napier a stabilit scara logaritmică, astfel cum se vede în fig. 2.2, în care apar diviziunile egale însemnate cu compasul. Scara construită cu rigla și compasul a constituit astfel primul instrument tehnic de calcul. Întrunirea a trei, respectiv a două scări divizate logaritmic, a condus la construirea tablei de înmulțire cu trei scări, precum și a riglei de calcul. [5]

În anul 1623, Schickard a construit pentru prietenul său, matematicianul și astronomul Johannes Kepler, un ceas de calcul bazat pe principiile primelor mașini de calcul de birou, iar filozoful francez Blaise Pascal a realizat în 1642 un dispozitiv de numărare cifric, prevăzut cu un șir de roți dințate. Cu treizeci de ani mai tîrziu, filozoful și matematicianul german Gottfried Wilhelm Leibniz a construit o mașină de calculat, bazată pe principiul înmulțirii prin adunare repetată. Această mașină a fost perfecționată de Charles Xavier Thomas în 1820 și ea este considerată precursoarea mașinilor de birou, îmbunătățite continuu pînă în zilele noastre.

În 1920 au fost introduse mașinile de calcul *electromecanice*, care au fost apoi perfecționate prin introducerea sistemului cu cartele perforate, Dealtfel, acest sistem datează încă din primul deceniu al secolului al XVIII-lea, cînd a fost aplicat în Franța, în tehnologia războaielor de țesut (în special de J. M. Jacquard). Matematicianul englez Charles Babage a încercat în 1823 construirea unei mașini *(Difference Engine)* destinate calculului funcțiilor, iar în 1829 a alcătuit proiectul pentru *Analitical Engine*, prima încercare de a construi, cu elemente mecanice, o mașină de calcul condusă pe bază de programe și memorie. El a folosit întîia oară cartele perforate într-o operație matematică. În 1887, Hermann Hollerith a realizat un sistem mecanic de înregistrare, compilare și prelucrare a datelor referitoare la recensăminte, cu utilizarea *cartelelor perforate*.

Cartela perforată este un purtător material de informație constînd dintr-o fișă de carton standardizată (la dimensiunile: lungimea 187,4 $\pm$ 0,1 mm; lățime 82,5 $\pm$ 0,1 mm; grosime care variază între 0,18—0,015—0,020 mm), cuprinzînd 45 sau 80 de coloane și 12 poziții sau rînduri, pe care se înregistrează mecanic datele din documentele primare sau din foile de codificare, perforîndu-se cartela cu ajutorul unei machete sau scheme și a unei mașini. Obișnuit, aceasta se compune dintr-un mecanism de alimentare și evacuare a cartelelor,

[5] N. KÖRWIEH, *350 Jahre Rechentechnik*, în „Rechentechnik-Datenverarbeitung" nr. 4/1972, Berlin.

un mecanism de culegere și un mecanism de perforare. Cu o mașină similară se efectuează operațiunea de verificare a transmiterii exacte pe cartelă a datelor. Operația de *sortare*, respectiv de grupare a cartelelor, se desfășoară, de asemenea, cu ajutorul unei mașini, care în cazul în care este înzestrată cu contori poate executa și numărarea perforațiilor din una sau mai multe coloane ale cartelei. La aceste operații și dispozitive trebuie să adăugăm *tabularea*, adică operația de imprimare în *tabulogramă* a datelor de identificare și grupare a indicatorilor, de control al grupelor, de calcul și imprimare a datelor individuale și totale, cu ajutorul mașinii de tabulat.

Mijloacele mecanice arătate sînt grupate respectîndu-se o anumită proporționalitate, de pildă o mașină de tabulat, o mașină de sortat, trei mașini de perforat și două mașini de control al perforării, alcătuind ceea ce se numește un set de mașini de calcul cu cartele perforate. Acesta poate fi completat și cu alte mașini, în scopul sporirii randamentului.

Tehnica automată de tratare a informației. Apariția calculatorului electronic și constituirea informaticii ca știință și tehnică a prelucrării automate au revoluționat sistemul informațional. A fost acoperită lipsa unui mijloc rapid de culegere, prelucrare, tezaurizare și transmitere a informației și a devenit posibilă colectarea, de către un singur organ, a tuturor datelor provenite din diverse compartimente ale unui organism (întreprindere, centrală ș. a.), tratarea datelor primare în scopul obținerii de informații derivate, difuzarea informațiilor pe activități, efectuarea de operații succesive sau independente și coordonarea în cadrul unui proces automat a funcțiilor interconectate ale sistemelor din natură și societate.

Calculatorul automat îl eliberează pe om de muncile de rutină și migăloase de calcul, în care erorile apar frecvent; el sporește valorificarea inteligenței umane, în sens extensiv, degrevîndu-l pe om de munca fizică și de partea muncii intelectuale susceptibilă de tipizare, în sens intensiv, amplificînd creativitatea prin analize aprofundate, prin meticulozitate și previziuni fundamentate. Tratarea automată a informației aproape că anulează factorul timp. Calculatorul lucrează după un program gata întocmit, astfel că răspunsul său este dat aproape instantaneu. Descrierea formală a procesului de prelucrare automată a informației a înlesnit studiul mecanic al sistemelor de reglare automată a diverselor procese. Dar poate cea mai importantă caracteristică a acestei epocale realizări științifice și tehnice rezultă din faptul că ordinatorul îl obligă pe om să gîndească precis și limpede ce anume urmărește și ce trebuie să facă pentru a atinge obiectivul urmărit. Aici apare și se concretizează rolul excepțional al calculatorului automat în conducerea activităților complexe umane.

De la încercarea nereușită a lui Ch. Babage, din 1829, de a construi cu elemente mecanice o mașină de calculat condusă pe bază de program a trebuit să treacă un secol pînă la elaborarea, în 1939, de către dr. Howard Aiken de la Universitatea Harvard (S.U.A.), a planurilor unei mașini numite calculator automat cu cicluri comandate *(Automatic Sequence Controlled Calculator)*. Este vorba de o dezvoltare a mașinii de perforat cartele, construită de International Business Machines Corporation (I.B.M.) pe principiul calculatorului lui H. Hollerith. Dar încă în 1934, Konrad Zuse începe pregătirile pentru un calculator cu releuri, folosind pentru prima dată sistemul numeric binar. K. Zuse a preconizat și folosirea tuburilor electronice, precum și un sistem de calcul pentru programare. Proiectele sale însă nu s-au bucurat de interes. Abia în 1944 a fost pus în funcțiune calculatorul cu releu MARK I, construit pe baza proiectului lui H. Aiken

Scanned with OKEN Scanner

pentru mașina de perforat I.B.M. și combinînd într-un singur dispozitiv capacitatea de a primi datele la intrare, de a realiza o succesiune de calcule și de a înregistra rezultatele la ieșire. Adevăratul calculator electronic, numit prescurtat ENIAC *(Electronic Numerical Integrator and Computer* = integrator și calculator electronic numeric), a fost realizat în 1946 de I.P. Eckert și J. W. Mauchly, de la Universitatea din Pennsylvania (S.U.A.). La acest calculator au fost utilizate pentru prima dată tuburi electronice, în locul dispozitivelor mecanice sau a comutatoarelor electromecanice; el a fost înzestrat cu capacitatea de a rezolva probleme pe baze logice. Calculatorul UNIVAC *(Universal Automatic Computer)*, livrat biroului de recensăminte în 1951, a fost utilizat aproape continuu 24 de ore pe zi, cîte 7 zile pe săptămînă, timp de peste 12 ani, iar în 1963 „s-a considerat că ar prezenta un interes istoric suficient pentru a fi așezat în expoziția de la Institutul Smithsonian".

Dar primele echipamente electronice de calcul erau încă niște „monștri greoi", cîntărind circa 30 de tone, ca ENIAC, cuprinzînd 17468 tuburi, consumînd pentru funcționare cam 174 kW curent și necesitînd încă mai mult pentru instalația de climatizare. Tehnica de construcție și performanțele calculatoarelor s-au dezvoltat rapid și sînt caracterizate prin „generații".

Prima generație (1951—1953), reprezentată de calculatoare cu tuburi electronice, a asigurat o accelerare considerabilă și exactitate în calculele științifice, tehnice și, într-o oarecare măsură, în cele economice. Calculele rapide se puteau efectua după un procedeu logic clar, dar calculatoarele din această generație nu puteau face față unui volum mare de date.

A doua generație (1958—1960) se caracterizează prin utilizarea tranzistorilor cu elemente logice active. Ea deschide era prelucrării unei mari cantități de date, posibilități dezvoltate de calculatoarele din cea de a treia generație (1963—1965). Prin crearea de circuite tranzistorizate, construirea primelor sisteme operaționale cu un număr de limbaje, multiprograme, programe-model, crește considerabil posibilitatea utilizării calculatoarelor de către diverse categorii de beneficiari.

În anii 1969—1972 se construiesc calculatoare din generația a patra, care folosesc ca elemente logice active ansambluri funcționale integrate de mărime medie, iar în *software*[6] limbaje cu posibilități multiple, programe metacompilate în *hardware*[7], sisteme pentru dialogul om-calculator. Memoria se bazează pe semiconductori, calculatoarele sînt combinate în serii mari și se construiesc în serie minicalculatoare. Sistemele sînt apte să răspundă de la stații îndepărtate, primind denumirea de *on-line* și bucurîndu-se de proprietatea *real-time*, ceea ce înseamnă că datele sînt prelucrate atît de repede încît pot fi utilizate instantaneu, iar calculatorul poate servi concomitent mai mulți solicitanți.

De pe acum este deschisă calea pentru cea de a cincea generație, care-i va asocia omului, definitiv, prin calculator, un auxiliar de mare precizie, rapiditate și eficiență. Datorită noilor înzestrări tehnice omul va beneficia de ansambluri funcționale integrate și memorii de mari dimensiuni. Datorită utilizării circuitelor cu un grad tot mai ridicat de integrare și reducerii dimensiunii fizice a aparatelor, se reduc cheltuielile privind latura tehnică, crește numărul funcțiilor din sistem, iar memorizarea devine tot mai rapidă. Se simplifică accesul la calculator și se generalizează sistemele de lectură automată a documentelor. Debitul imprimatoarelor, ca și viteza de transmitere se multiplică, reducîndu-se totodată decalajul dintre timpul necesar pentru introducerea informațiilor în calculator și utilizarea informațiilor elaborate. Adăugarea unor unități terminale și programe modulare corespunzătoare permit un dialog cu calculatorul, în sensul adevărat al cuvîntului.

Se obișnuiește a se clasifica calculatoarele electronice astfel: a) după destinație: calculatoare specializate și universale; b) după tipul, adică după modul în care sînt prezentate datele: analogice și numerice; c) după capacitate,

[6] Ansamblu de activități precum: codificarea, organizarea, analiza, programarea.
[7] În opoziție cu *software*, ansamblu de elemente constituind echipamentul tehnic electronic.

Scanned with OKEN Scanner

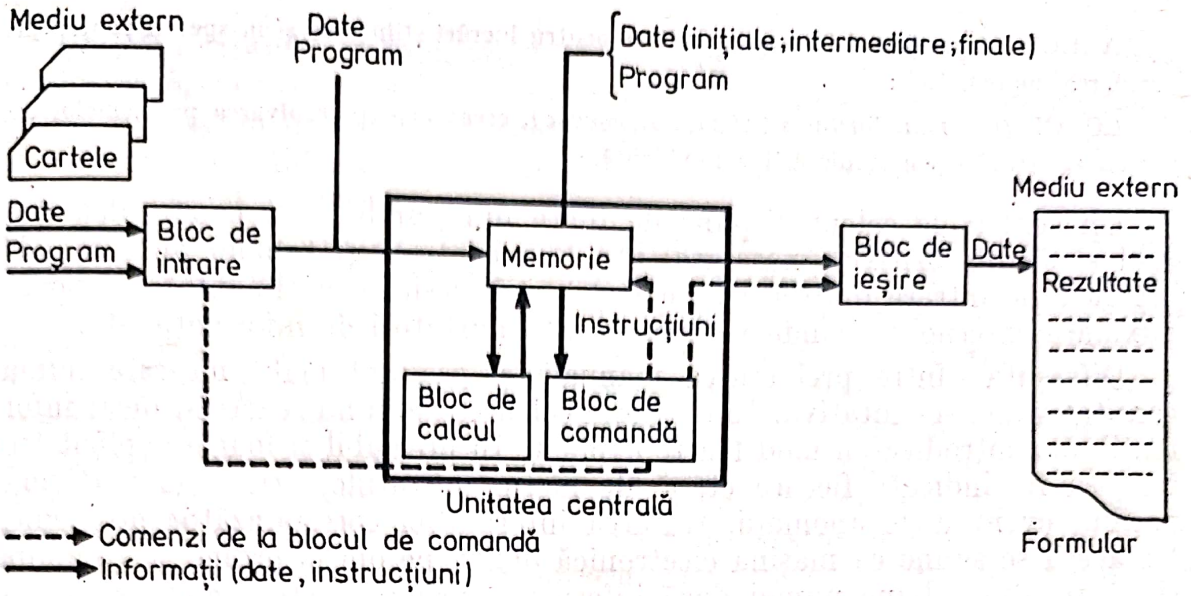


Fig. 2.3 Structura calculatorului electronic [68]

respectiv în funcție de volumul de operații posibile de executat, viteza de lucru, volumul de date pe care dispozitivele componente ale calculatorului le pot memora și prelucra, caracteristici care constituie elemente de diferențiere a generațiilor.

Structura calculatorului este formată din blocurile de intrare, de ieșire, de memorie, de calcul și de comandă, care schematic pot fi prezentate ca în fig. 2.3.

Pentru funcționarea calculatorului, un organ de memorie conține concomitent și definirea detailată a etapelor de calcul, datele și zonele de lucru necesare. Informațiile sînt introduse în memorie sub formă codificată. Un organ de comandă examinează una cîte una etapele de calcul, numite instrucții de program și localizează în memorie factorii fiecărei operații. Fiecare instrucție este exercitată prin circuite corespunzătoare, ansamblul lor constituind unitatea de calcul. Prin comparație, memoria calculatorului este analoagă foii de hîrtie sau fișierului din sistemul manual de prelucrare a informației, programul este analogul suportului de calcul, unitatea de comandă este analogul angajatului (funcționarului) și unitatea de calcul — analogul mașinii. Organele de intrare și ieșire au rolul de a transfera informațiile din mediul exterior în memoria calculatorului și invers [68].

Potrivit schemei generale de transmitere a informației, calculatorul primește informația pentru tratare, prin codificare sistematizată, numită *limbaj*, iar succesiunile de operații ce trebuie să le execute pentru rezolvarea problemei sînt transpuse în *program*.

Limbajele de programare a instrucțiunilor date calculatorului, foarte diferite de descrierea matematică uzuală a desfășurării operațiilor de calcul, sînt de mai multe feluri: limbaje mașină, alcătuite din instrucțiuni specifice construcției calculatorului, limbaje simbolice, care folosesc mnemonici (prescurtări ale numelui operației), limbaje nealgoritmice, care nu sînt utilizate pentru a descrie algoritmi, limbaje convenționale sau interactive, utilizate de pildă în cercetare sau în procesul de învățămînt, în care nu pot fi prevăzute detaliat toate posibilitățile de calcul [68]. Limbajele algoritmice mai des folosite sînt:

FORTRAN *(Formula Translator)*, în care se pot efectua cu ușurință prelucrări cu un volum mare de calcule, folosit pentru programarea lucrărilor cu caracter științific;

Scanned with OKEN Scanner

ALGOL *(Algoritmic Language)*, utilizat pentru lucrări științifice și în special pentru programarea matematică;

COBOL *(Commun Business Oriented Language)*, creat pentru rezolvarea problemelor economice și mai ales pentru lucrările de evidență. [8]

Un centru de calcul dispune de toate funcțiunile unui mare calculator, deci și de dispozitivele necesare: unitatea centrală și tabloul de comandă, organele de intrare (cititorul de cartele) și de ieșire (imprimanta), memoriile auxiliare externe pe bandă magnetică, alți purtători de informație etc.

Diferența dintre prelucrarea manuală și cea automată nu este numai cantitativă, ci și calitativă. În memoria calculatorului nu se găsesc decît informații, dar introduse în mod foarte explicit. În prealabil și foarte explicit trebuie să fie indicate fiecare etapă de lucru, operațiile, orice caz particular posibil, eventualele anomalii, tratarea informației corespunzător problemei. De aceea se spune că mașina electronică obligă pe om la ordine și disciplină și dă rezultate bune numai dacă informația primară este corectă și îi este transmisă foarte detaliat și precis. Calculatorul nu rezolvă probleme în general, ci se ocupă numai cu ceea ce i se prescrie prin program.

Calculatoarele electronice sînt puse în situația de a participa direct la dirijarea activităților (proceselor), indiferent de natura acestora, de gîndire sau producție materială, de cercetare sau învățare. Participarea capătă caracterul unui dialog, unei conversații om-mașină, constînd în esență în stabilirea unei comunicații între două sisteme total diferite. Unul — omul —, capabil de gîndire creatoare, poate lua decizii și în condiții de incertitudine, dar citește, calculează și scrie lent și este predispus la erori. Celălalt — mașina — are o putere asociativă limitată, dar calculează, citește și scrie rapid și greșește foarte rar. În dialogul om-calculator aceste două sisteme sînt reciproc dependente, urmărind un anumit scop, printr-o acțiune comună.

Sistemul informațional statistic în cadrul conducerii cu prelucrarea automată a datelor. Norbert Wiener a demonstrat pentru prima dată că cel mai complex proces de conducere social-economică poate fi studiat cu ajutorul metodelor matematice și al altor științe, ca un proces tehnologic al prelucrării informației. Cercetarea operațională, definită foarte succint drept „instrument matematic al științei conducerii" [9], introduce în procesul de perfecționare a activității utilizarea modelului matematic și a unei game variate de metode statistice și matematice, în scopul optimizării deciziilor de conducere.

Între cibernetică, cercetarea operațională și prelucrarea electronică a informației este o legătură foarte strînsă. Cînd se face apel la cercetarea operațională, pentru a găsi varianta optimă a suportului rațional de decizie, obiectul acesteia este tratat ca element al unui sistem cibernetic, pentru cunoașterea concomitentă atît a procesului de ansamblu, cît și a proceselor parțiale, care caracterizează acel obiect. Legătura dintre cercetarea operațională și tratarea electronică a informației se manifestă în faptul că prelucrarea electronică formează baza tehnică a cercetării operaționale.

[8] La calculatoarele din țara noastră se folosesc, de pildă, limbajele denumite ASIRIS, pentru calculatorul FELIX C 256, ASSEMBLER, pentru IBM 360 etc.

[9] T. POPOVICI, *Metode și mijloace matematice moderne ale conducerii științifice*, în vol. *Conducerea științifică a întreprinderii*, Editura politică, București, 1970.

Scanned with OKEN Scanner

Datorită faptului că folosirea calculatorului electronic face posibilă automatizarea prelucrării aproape a tuturor datelor din compartimentele de activitate, el devine un adevărat centru operațional.

Crearea sistemului de conducere cu prelucrarea automată a datelor este obiectivul final al perfecționării sistemului informațional social-economic; prin el se asigură interacțiunea necesară dintre: a) mecanismul operațional de realizare a producției de bunuri și servicii, ca obiect al activității sistemului economic; b) factorii de conducere, ca organ decizional pentru înfăptuirea obiectivelor sistemului economic; c) fluxul informațional, factorul de legătură dintre decident și organul operațional; fiecare, la rîndul său, poate fi considerat un sistem. Acest obiectiv poate fi atins numai în cursul unui proces îndelungat de reorganizare și adaptare a sistemului de conducere al întreprinderii, a metodelor de lucru, a fluxurilor de informație, a formelor de prezentare a datelor și rezultatelor prelucrării acestora. La capătul acestui proces, *sistemul informațional devine sistem informatic integrat.* Sistemul de conducere cu mijloace de prelucrare automată a datelor asigură, prin mijloacele și procedeele sale, legătura dintre elementele constitutive ale activităților de producție și cele de organizare și conducere. Între acestea din urmă și procesul de prelucrare a informațiilor se stabilesc în mod riguros anumite relații. Se creează astfel o succesiune de procese (decizie — prelucrarea datelor — producție / activități — prelucrarea datelor — decizie), care stau la baza funcționării organismului economic respectiv.

Introducerea sistemului de conducere cu mijloace de automatizare, în toate verigile economiei noastre naționale, constituie o direcție de acționare, care răspunde prevederii formulate în Programul Partidului Comunist Român de făurire a societății socialiste multilateral dezvoltate și înaintare a României spre comunism: „promovarea largă în munca de organizare și conducere a societății — inclusiv a proceselor și fenomenelor sociale — a metodelor moderne de analiză, decizie și prelucrare automată a datelor" [2]. Potrivit programului trasat de Hotărîrea Comitetului Central al Partidului Comunist Român, din 18 aprilie 1972, în țara noastră se va realiza sistemul de conducere cu prelucrarea automată a datelor, în scopul înfăptuirii unei conduceri științifice eficiente la toate nivelele, „în unitățile economice, pe ramuri și funcționale la nivelul economiei și integrarea lor ulterioară într-un sistem unitar național de prelucrare automată a datelor" [3].

Introducerea automatizării prelucrării datelor din sistemul informațional al conducerii social-economice nu afectează cu nimic funcția statisticii. În cadrul acestui sistem, statistica instituțională vehiculează o mare parte din informația necesară fundamentării deciziilor, elaborării planurilor de dezvoltare și urmăririi realizării acestora, cunoașterii și caracterizării tendințelor și perspectivelor în dezvoltarea multilaterală a societății.

Alimentat de progresele științei în general și ale statisticii, în particular, sub incidența noilor orizonturi ale informaticii, sistemul informațional statistic se încadrează în acțiunile generale de perfecționare a organizării, conducerii și planificării economico-sociale, a metodelor de organizare a muncii, corespunzător cerințelor vieții și schimbărilor continue ce au loc în societate. În acest sens, perfecționarea sistemului informațional statistic este canalizată spre o serie de obiective, precum: îmbunătățirea sistemului de indicatori statistici, definirea lor unitară, perfecționarea metodologiilor de calcul, îmbunătățirea normelor metodologice și a surselor de date, perfec-

Scanned with OKEN Scanner

Fig. 2.4 Schema

ționarea și dezvoltarea sistemului unitar de clasificări și nomenclatoare economico-sociale, valorificarea superioară a informațiilor statistice intrate în sistem, îmbunătățirea operativității informațiilor statistice și raționalizarea fluxurilor lor, ierarhizarea volumului informațiilor corespunzător cu atribuțiile și competențele fiecărei trepte organizatorice, perfecționarea purtătorilor de informație și crearea băncii centrale de date statistice [23].

Crearea sistemelor informatice funcționale presupune grefarea pe aceste sisteme a băncilor de date, precum și constituirea unei bănci centrale de date. Concepută ca un ansamblu de date cu privire la un anumit domeniu și depozitate sub forma unui sistem tehnic-organizatoric, *banca de date* este conectată la o rețea de *baze de date*, acestea fiind definite ca ansambluri structurate de fișiere. Prin constituirea băncilor de date se urmărește accelerarea proceselor de colectare, prelucrare și valorificare a datelor, eliminarea înregistrărilor repetate, asigurarea unui acces rapid la datele înmagazinate și selecționarea lor potrivit unor criterii stabilite în prealabil, automatizarea informării. Ca atare, organizarea băncilor de date presupune posibilitatea prelucrării automate a datelor, tezaurizarea de programe, relații directe între utilizatori și bancă.

Tabelul statistic și graficul statistic. Indiferent dacă prelucrarea este manuală sau, dimpotrivă, automată, rezultatele ei trebuie să fie înfățișate într-o formă care să le facă proprii utilizării de către beneficiar, să înlesnească aplicarea metodelor statisticii descriptive și, mai departe, analizei. Două procedee s-au impus în decursul istoriei statisticii: tabelul și graficul, avînd proprietăți nu numai de descriere, ci și de analiză.

Într-o reprezentare simplistă și evident unilaterală, tabelul statistic este cea mai plastică reprezentare a statisticii. În realitate însă, chiar cel mai elementar tabel presupune o operație de ordonare logică, de sistematizare care conferă materialului statistic posibilități de reflectare a unor situații sau procese reale. Același lucru se poate spune despre grafic. Într-o formă ilustrativă, pot fi înfățișate legături, tendințe, structuri, care înlesnesc cunoașterea și previziunea.

Desigur, tipologia tabelelor și graficelor este vastă și va putea fi cunoscută de-a lungul capitolelor următoare. În cele ce urmează, se dau numai unele elemente generale.

Un *tabel statistic* bine alcătuit dispensează de pagini de comentarii și interpretări, fiind un instrument de lucru deosebit de util pentru specialiști. Tehnica de prezentare a datelor în forma tabelară a devenit atît de răspîndită încît s-a ajuns la întocmirea de lucrări compuse numai din tabele, cele mai cunoscute și uzitate fiind anuarele statistice și publicațiile periodice ale diverselor oficii statistice naționale și internaționale.

Din punctul de vedere al conținutului, tabelul statistic este o culegere de cifre astfel ordonate încît să prezinte simplu și clar desfășurarea unui proces, repartiția unei colectivități sau interdependența dintre anumite fenomene.

Tehnica tabelară a evoluat în timp, stabilindu-se anumite principii de întocmire și elementele strict necesare. Forma tabelului statistic este dreptunghiulară, iar datele cifrice sînt înscrise astfel încît să formeze șiruri concomitent verticale (coloane) și orizontale (rînduri).

Scanned with OKEN Scanner

Orice tabel statistic cuprinde un subiect, care indică obiectul observării statistice, de obicei în rînduri, precum și un predicat, adică ceea ce se spune despre subiect (ca la analiza gramaticală), în coloanele tabelului.

După modul de prelucrare a datelor observării, tabelele statistice pot fi simple sau combinate.

Tabelul statistic simplu prezintă o serie statistică rezultată din prelucrarea datelor în raport cu o singură caracteristică (atributivă, de timp sau spațiu). Macheta acestui tabel este formată din două coloane: prima destinată înscrierii aspectelor caracteristicii după care s-a constituit seria statistică, iar a doua, înscrierii efectivului sau nivelului fiecărui aspect.

Tabelul statistic combinat prezintă serii statistice rezultate din gruparea simultană a unităților de observare după două sau mai multe caracteristici sau variabile. Forma cea mai tipică a unui asemenea tabel cu „dublă intrare" este *tabelul de corelație*, care înfățișează serii de repartiție în raport cu două variabile legate cauzal sau interdependente.

Reprezentările grafice înlocuiesc cifrele prin combinații de linii sau prin figuri geometrice; ele transmit deci informația statistică vizual. William Playfair, părintele tehnicii grafice, aprecia că prin această metodă se pot imprima în memorie, în cinci minute și durabil, informații a căror culegere și transpunere în tabele ar dura zile întregi. Autorul francez André Piatier consideră că o reprezentare grafică simplă și clară constituie un suport al gîndirii.

Orice mărime poate fi reprezentată printr-un număr, iar geometric, printr-un segment, valoarea absolută a numărului fiind exprimată prin lungimea segmentului. Alegerea lungimii segmentelor care reprezintă unitățile de măsură este arbitrară. Dacă pe axa XX' (fig. 2.5) alegem un punct fix O (originea), fiecărui număr real x îi va corespunde un anumit x pe axă; segmentul OA al axei va fi măsurat prin x.

Fig. 2.5 Reprezentarea unei mărimi pe o axă.

Analog se poate determina poziția unui punct în plan, și anume prin metoda coordonatelor, din care cele mai des folosite sînt coordonatele rectangulare (coordonatele lui Descartes sau carteziene). Dacă ducem în plan două drepte perpendiculare XX' și YY' (fig. 2.6), numite axe de coordonate și considerăm ca origine punctul lor de intersecție O, oricărui punct din plan îi corespund

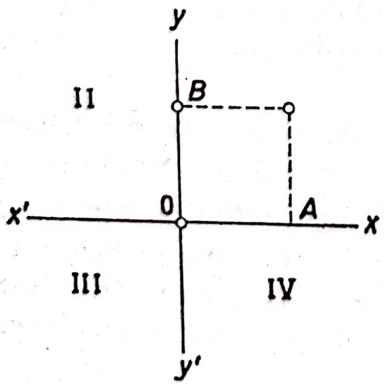


Fig. 2.6 Determinarea unui punct în plan.

Scanned with OKEN Scanner

două numere reale, care măsoară segmentele determinate de proiecțiile punctului pe axele coordonate (axa XX' și axa YY').

Reciproc, fiecărei perechi de numere (x, y) îi corespunde un punct în plan.

Pe axa XX' se ia un segment de mărimea x, numită abscisa (punctul A), pe axa YY' se ia un segment de mărimea y, numită ordonata (punctul B). Ducînd paralele prin punctele C și B la axele coordonate, se obține punctul M (x, y) la intersecția acestor paralele. Axele de coordonate împart planul în 4 cadrane, numerotate I, II, III și IV, în sens invers mișcării acelor ceasornicului. Semnele coordonatelor punctelor din cele 4 cadrane sînt date mai jos:

M	I	II	III	IV
Abscisa $= x$	+	−	−	+
Ordonata $= y$	+	+	−	−

OA — abscisa $= x$
OB — ordonata $= y$

Metoda coordonatelor se folosește în statistică pentru reprezentarea grafică a repartițiilor. Valorile variației independente sînt reprezentate prin puncte pe axa absciselor și frecvențele prin puncte pe axa ordonatelor. Se pot obține trei tipuri de reprezentări grafice: histograma, poligonul frecvențelor și curba frecvențelor.

Orice funcție $y = f(x)$ poate fi reprezentată grafic; dînd lui x diverse valori, se calculează valorile corespunzătoare ale lui y. Fixînd în plan cît mai multe puncte cu coordonatele astfel determinate și unindu-le printr-o linie continuă, obținem o curbă care este reprezentarea grafică a funcției $y = f(x)$.

Reprezentarea grafică este o construcție de ansamblu, în care se disting anumite elemente de bază, comune și specifice. De obicei, o reprezentare grafică cuprinde: titlul, scara de reprezentare, rețeaua, semne convenționale și explicarea lor, note, construcția propriu-zisă (figura).

Titlul este un element constitutiv ca și la tabelele statistice și trebuie să fie scurt, clar și complet, adică din titlu trebuie să rezulte obiectul reprezentării grafice, timpul și spațiul la care se referă datele reprezentate.

Scara de reprezentare, care este o linie ale cărei puncte pot fi citite ca numere precis determinate, apare sub diferite forme și combinații și se compune dintr-un suport (linie) și puncte numerotate pe acesta. Scara unei variabile x este locul geometric al punctelor cotate, corespunzătoare unei valori determinate a variabilei x. Cu ajutorul scării de reprezentare, utilizînd un punct de pe suport, se găsește valoarea variabilei sau — dacă aceasta este cunoscută — se identifică punctul respectiv de pe scară. În felul acesta se transformă în dimensiuni mărimea indicatorilor, care caracterizează fenomenul sau procesul urmărit.

Scanned with OKEN Scanner

Scara este rectilinie sau curbilinie, după cum este constituit suportul: dintr-o linie dreaptă sau curbă. Ea este uniformă sau neuniformă, în funcție de caracterul egal sau inegal al intervalelor grafice, adică al segmentelor dintre punctele cotate sau grupele de puncte. Exemplu tipic de scară uniformă este scara aritmetică, utilizată la reprezentarea repartițiilor statistice cu variabile cantitative continue, între care există aproximativ o legătură funcțională liniară. Scări neuniforme sînt scara logaritmică, prin care se estimează o constantă 0 cu semnificație de variație relativă într-o relație de forma $y = 0$, scara gaussiană, pe care se reprezintă valorile abaterii normate u sau frecvența cumulată $\frac{n_i}{n}$, segmentele dintre diviziunile scării fiind egale în raport cu valorile variabilei normate u și neegale în raport cu frecvența cumulată a repartiției, precum și scările binomiale. Scările curbilinii foarte cunoscute sînt cele circulare (de ex. cadranul ceasului) și în arc (scările multor aparate de măsură).

În fig. 2.7, *a* este scara aritmetică (uniformă), iar *b* — scara geometrică (logaritmică).

Scările uniforme se construiesc prin utilizarea unor unități de măsură (milimetri sau altele) pentru fixarea punctelor pe suportul scării și notarea lor

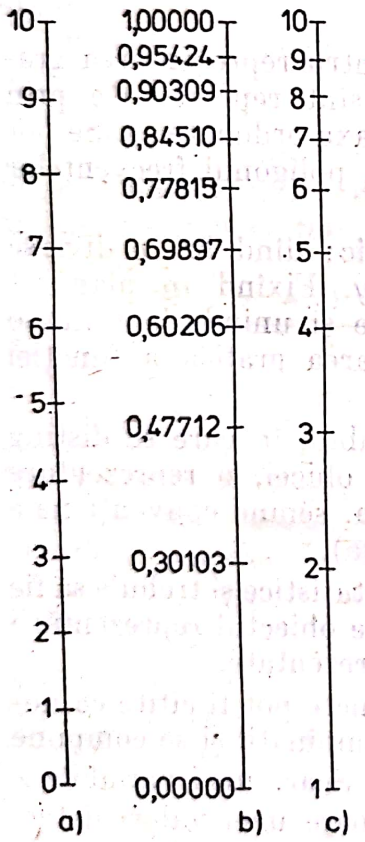


Fig. 2.7 Tipuri de scări verticale pentru reprezentări grafice: a) scară aritmetică; b) scară logaritmică cu diviziuni corespunzătoare antilogaritmilor; c) scară logaritmică cu diviziuni de la 1 la 10.

cu numere după specificul obiectului (fenomenului) reprezentat. La scările logaritmice distanța de la originea scării la punctele de pe suport este proporțională cu logaritmii numerelor, și nu cu mărimea acestora. Ca atare, procedeul cel mai simplu de construire a acestor scări constă în utilizarea tabelelor loga-

ritmice. Pentru ușurarea construirii scării logaritmice, în tabelul care urmează se dau logaritmii numerelor 1—10 și 10—100:

Numere	1	2	3	4	5	6	7	8	9	10
Logaritmi	0	0,30	0,48	0,60	0,70	0,78	0,85	0,90	0,95	1,00
Numere	10	20	30	40	50	60	70	80	80	100
Logaritmi	1	1,30	1,48	1,60	1,70	1,78	1,85	1,90	1,95	2

Dacă aplicăm logaritmii numerelor din tabelul de mai sus pe suportul scărilor vom constata că se repetă intervale absolut omogene între numerele 10—100 ca și între numerele 1—10.

De obicei, gradația scării figurează pe verticală la stînga reprezentării grafice, uneori și la dreapta, după cum convine mai bine, în funcție de locul plasării figurii. Se poate, de asemenea, ca fenomenul reprezentat să necesite utilizarea de scări atît la stînga, cît și la dreapta. Fig. 2.7 prezintă trei tipuri de scări de reprezentare pe verticală: *a* este scară aritmetică; *b* — scară logaritmică în care sînt menționați logaritmii; *c* — scară logaritmică pentru indicarea datelor problemei.

Rețeaua graficului constituie elementul de interpretare a reprezentării grafice și este formată din linii paralele orizontale, verticale sau încrucișate ori în cercuri concentrice, care ajută la determinarea valorii (mărimii) asociate diferitelor puncte.

O utilizare foarte largă în analiza proceselor social-economice o au rețelele rectangulare care se construiesc în sistemul coordonatelor rectangulare; de obicei, la construirea acestor rețele se folosește cadranul I din planul axelor de coordonate (și uneori cadranul IV împreună cu cadranul I), cum s-a văzut în figura 2.6.

O clasificare generală a reprezentărilor grafice reține următoarele tipuri:

a) după formă geometrică de reprezentare — grafice reprezentate prin linii, benzi, unghiuri, suprafețe, volume etc.;

b) din punctul de vedere al geometriei analitice — grafice cu coordonate rectangulare, grafice cu coordonate polare etc.;

c) din punct de vedere analitic — grafice cu scări uniforme, grafice cu scări funcționale;

d) după natura repartițiilor și a seriilor statistice care urmează să fie reprezentate grafic — histograme, poligoane de frecvență, curbe de frecvență (pentru repartițiile statistice cu o singură caracteristică), cartograme sau cartodiagrame (pentru repartițiile după o caracteristică teritorială) etc.;

e) după scopul sau destinația construcției grafice — diagrame de comparație, grafice de plan, armonograme, ciclograme, nomograme etc.

Diagrama prin benzi se utilizează pentru reprezentarea grafică, în formă geometrică dreptunghiulară, a unor repartiții cu caracteristici calitative, lungimea benzilor fiind direct proporțională cu datele reprezentate.

Diagrama prin coloane se deosebește de precedenta prin așezarea formei geometrice dreptunghiulare de reprezentare în sens vertical, înălțimea coloanelor fiind direct proporțională cu frecvența caracteristicilor reprezentate,

Scanned with OKEN Scanner

cu condiția ca, coloanele să aibă bazele egale. Construcțiile grafice cu coloane suprapuse poartă denumirea de diagrame cu coloane în aflux și ele se folosesc adesea în scop de popularizare. Nu există o regulă rigidă pentru ierarhizarea indicatorilor reprezentanți prin astfel de coloane.

Diagramele prin areale folosesc cel mai des, dintre figurile geometrice, pătratul, cercul și dreptunghiul. La aceste construcții aria figurii geometrice este proporțională cu frecvența modalității caracteristicii calitative ce formează obiectul reprezentării.

Diagrama de structură poate fi prezentată prin cercuri. La construirea unei asemenea diagrame se consideră că cercul, căruia îi corespund 360°, reprezintă totalitatea colectivității considerate, iar greutatea specifică a elementelor structurale se determină prin calcul, pe baza regulii de trei simplă.

Construcțiile grafice pentru reprezentarea repartițiilor cu o caracteristică cantitativă cu variația discretă sînt diagramele prin benzi sau linii, histogramele, poligoanele de frecvență și curbele de frecvență.

Cronograme (historiograme) se numesc construcțiile grafice tipice pentru seriile statistice dinamice, care pot fi momente sau intervale.

Cartogramele sau hărțile statistice sînt construcții grafice avînd la bază hărți obișnuite pe care se înregistrează grafic observările statistice astfel încît să fie asociate spațiului geografic la care se referă, constituind instrumente de reprezentare concretă a „poziției geografice a observărilor".

Stereogramele sînt diagrame de reprezentare, prin figuri geometrice de volum (paralelipipede), a unor fenomene caracterizate de patru indicatori componenți, cu condiția ca mărimea celei de a patra caracteristici să fie egală cu produsul celorlalte trei.

Grafice cu trei dimensiuni, bazate pe sistemul coordonatelor rectangulare în spațiu, se folosesc și pentru a reprezenta corelația a trei caracteristici.

Diagrama polară (radială), construcție grafică în formă de cerc, avînd la bază o rețea curbiliniară radială, se execută cu ajutorul coordonatelor polare, nu rectangulare și se folosește la reprezentarea unor fenomene cu variație sezonieră, lunară, săptămînală sau zilnică, dar care nu este prea pronunțată de la o perioadă la alta. Acest tip de diagramă se construiește prin reprezentarea grafică a indicatorilor seriei statistice fie prin segmente de dreaptă, fie prin sectoare de cerc.

Exemplificarea tipurilor și formelor de reprezentări grafice, mai des utilizate în practică, se găsește în cuprinsul capitolelor care urmează, secvențalor fiind adaptată profilului prezentei lucrări.

2.4 ANALIZA STATISTICĂ ȘI DEZVOLTAREA EI CA ANALIZĂ DE SISTEM

Considerată drept ultimă treaptă a cunoașterii statistice, analiza constituie o parte componentă a procesului de valorificare a materialului statistic obținut prin observare și în fazele prelucrării.

Noțiunea de analiză este utilizată în investigarea fenomenelor din natură și societate, pentru a desemna metoda de cercetare caracterizată prin descompunerea respectivelor fenomene în elemente constitutive, în scopul ridicării cunoaș-

Scanned with OKEN Scanner

terii lor de la singular la general, de la concret la abstract, de la ceea ce este dat nemijlocit la mijlocit, de la întîmplător la necesar, de la ceea ce este aparent în fenomene la legitatea lor [34]. Valorificarea materialului statistic constă în oglindirea și analiza fenomenelor de masă, corespunzător obiectivelor cercetării întreprinse și ca atare ea influențează modul de desfășurare a acesteia pe treptele anterioare. În procesul valorificării este apreciat întregul material rezultat și prezentat în diferite forme.

Aplicată în procesul valorificării statistice, metoda analizei pătrunde în esența fenomenelor, desăvîrșind desfășurarea cercetării statistice și desprinzînd concluziile teoretice și practice, pe baza datelor rezultate din observare și prelucrare. În acest scop se supun investigării metodice factorii sub a căror influență se desfășoară fenomenul studiat, se cercetează legitățile, cauzele, tendințele, legăturile și acțiunile reciproce din interconexiunile fenomenelor; metodele statistice la care se face apel trebuie să fie aplicate în mod diferențiat, ținîndu-se seamă de natura și modalitățile dezvoltării fenomenelor cercetate, de particularitățile acestora. Concluziile analizei trebuie să fie în concordanță cu datele realității și să aibă utilitate practică.

Analiza începe printr-o apreciere critică a desfășurării cercetării în etapele anterioare, urmărindu-se dacă au fost bine definite obiectul cercetării și procedeele de observare și dacă a fost tradus în întregime în viață programul observării și corect puse în aplicare procedeele de prelucrare. De asemenea, în prealabil, se examinează dacă în desfășurarea cercetării au apărut erori sistematice sau întîmplătoare, involuntare, din neglijență sau de altă natură.

În scopul cunoașterii aspectelor dinamice și interdependențelor din cîmpul fenomenelor cercetate, analiza se servește de comparații în timp, spațiu sau după natura caracteristicilor urmărite. În cazul unor fenomene similare se efectuează comparații statice, dinamice sau de raport între plan și realizare, iar dacă natura fenomenelor este diferită, se fac confruntări.

În cadrul mai larg, de pildă al analizei multifactoriale, avînd ca obiect cercetarea comportamentului statistic al mai multor caracteristici și urmărind stabilirea proprietăților repartițiilor multidimensionale, s-au dezvoltat diverse metode. Astfel, analiza dispersională, cunoscută și sub denumirea de analiza varianței, desfășoară o investigație calitativă a datelor de observare cu acțiune concomitentă, în scopul stabilirii celor mai importanți factori și a estimării influenței lor. Pe de altă parte, analiza factorială poate fi concepută ca o metodă de analiză globală și are ca obiect reprezentarea unui grup de variabile în funcție de cîțiva factori, care pot fi comuni (acționînd asupra mai multor variabile) și specifici (cu acțiune asupra unei singure variabile). Cu alte cuvinte, această metodă urmărește să explice cauzele de variație ale unui complex de variabile aleatoare (sau criterii), cu ajutorul unui ansamblu de variabile ipotetice independente (sau factori).

Informatica și-a pus amprenta asupra metodei statistice prin *analiza de sistem*, cu ajutorul căreia cunoașterea statistică pătrunde, pe de o parte, în relațiile sistemelor (fenomenelor) studiate cu mediul, pe de altă parte, în problematica lor internă, dezvăluind stările, structura, interconexiunile, conexiunile inverse care acționează în cadrul lor de manifestare. Raționalitatea și veridicitatea reflectării acestor aspecte de indicatorii statistici ai realității social-economice pot fi temeinic cercetate cu ajutorul analizei de sistem, considerînd că esența conceptului de sistem permite utilizarea lui nu numai în sfera realității materiale, ci și în operațiile intelectului. Cu ajutorul acestui

Scanned with OKEN Scanner

concept se delimitează obiectul abordării problemelor, fără ca totuşi să se pună condiţii marginale. Pentru statistică „se desprind cel puţin trei concluzii în legătură cu analiza de sisteme informaţionale: 1) să se gîndească pe sine ca sistem, prin metoda riguroasă a analizei de sistem informaţional; 2) să-şi îmbunătăţească procedeele de observare prin folosirea analizei de sistem sau prin preluarea, din analizele de sisteme informaţionale, a informaţiilor necesare statisticii; 3) să facă efortul de a înţelege fenomenele şi procesele social-economice ca sisteme sau făcînd parte din sisteme". [10]

Analiza de sistem poate ridica nivelul primelor etape ale cunoaşterii statistice, perfecţionînd programele observărilor statistice în direcţia asigurării informaţiilor necesare şi suficiente şi solicitînd sprijinul cercetărilor sociologice, mai ales al metodei anchetelor. Dacă pe această cale intervenţia analizei de sistem este puternic operantă la nivel microeconomic, aplicarea la nivel macroeconomic necesită abordarea problemelor şi „din perspectiva altor discipline, pe care le includem în noţiunea largă de ştiinţă a conducerii şi în care statistica are un cuvînt greu de spus, în special în determinarea conţinutului şi metodologiei indicatorilor." [11]

Aşadar, şi sub acest unghi de vedere, problema calităţii, a expresivităţii indicatorilor statistici social-economici se situează pe primul plan. Perfecţionarea lor vizează, în concepţia statisticienilor români [24], obiective ca: integrarea indicatorilor şi metodologiilor într-un ansamblu corelat, ierarhizarea lor pe trepte organizatorice, în conformitate cu atribuţiile şi competenţele organelor lor de conducere, extinderea principiului de informare după metoda abaterilor sau a excepţiei, codificarea sistemului de indicatori şi adaptarea lui la prelucrarea automată, eliminarea din sistem a indicatorilor sau detalierilor nesemnificative.

[10] M. BIJI, *Probleme ale statisticii economice desprinse din acţiunea de perfecţionare a sistemului informaţional din R.S. România*, referat prezentat la sesiunea ştiinţifică a Academiei de Studii Economice, Bucureşti, 1973.

[11] *Ibidem.*

Scanned with OKEN Scanner

CAPITOLUL 3
Probabilitate și variabilă aleatoare. Repartiții teoretice

Rezultatul unei observări statistice, al unui recensământ al populației, al unei înregistrări a utilajelor, al unei anchete asupra cererii populației, al unei experimentări agrobiologice, se centralizează și se prelucrează, căpătînd forma obișnuită a unei serii statistice. Obținem astfel repartiția unui număr de muncitori după procentul de îndeplinire a normelor, a numărului căsătoriilor după vîrsta soților, a pieselor după dimensiunea lor etc. Ele permit determinarea unor valori sintetice din care cea mai frecventă este media. În toate cazurile, statistic vorbind, avem populații finite: un număr oarecare de muncitori, un număr strict determinat de piese etc. Într-un fel vom raționa dacă ne bazăm pe rezultatele înregistrării căsătoriilor într-un oraș oarecare în ce privește vîrsta soților și în alt mod dacă vom avea toate căsătoriile din România în decursul unui an. Desigur, vom putea face ipoteze cu privire la valabilitatea sau reprezentativitatea unor asemenea valori, în funcție de mărimea populației statistice.

Din demersul modern al unei științe se știe bine că etapele sînt: teoria, modelul și testarea, care reproduce, în general, raportul dintre teorie și practică. În această optică, datele statistice sînt date experimentale sau empirice; ele se încadrează într-un model sau într-o lege sau, dimpotrivă, se „abat" de la acest model. Repartiția statistică poate fi văzută deci ca repartiție *empirică*, în raport cu o alta care este *teoretică*. În mod corespunzător vom distinge valori empirice și valori teoretice.

Conceptele care fac legătura între cele două tipuri de repartiții sînt cel de probabilitate și de variabilă aleatoare. S-ar fi putut porni de la empiric la teoretic; s-a preferat totuși cealaltă cale, adică drumul de la model la realitate, de la teoretic la empiric, prefațîndu-l cu noțiunea de probabilitate. La rîndul său, aceasta va fi prezentată într-o accepție clasică, apropiată de reprezentarea intuitivă, și una riguros științifică, cea axiomatică.

Pentru unele elemente de detaliu sau de demonstrație, cititorul este invitat să se adreseze lucrărilor de specialitate.

Scanned with OKEN Scanner

3.1 TEORIA PROBABILITĂȚILOR. CONCEPTE

Teoria probabilităților constituie una din cele mai importante ramuri ale matematicii contemporane. Modelele abstracte ale teoriei probabilităților sînt utilizate, deopotrivă, în tehnică, în fizică și chimie, în biologie și medicină, în astronomie, meteorologie, arheologie, lingvistică etc., ca și în studiul fenomenelor social-economice, al proceselor industriale moderne.

Obiectul de studiu al teoriei probabilităților îl formează legile ce se manifestă în domeniul fenomenelor întîmplătoare de masă. În sens larg, prin noțiunea de fenomen întîmplător vom înțelege acel fenomen care — la efectuarea repetată a unui experiment — se manifestă, de fiecare dată, într-un mod mai mult sau mai puțin diferit.

De pildă, cîntărind de mai multe ori unul și același corp, cu ajutorul *aceluiași* aparat de măsură, înregistrăm de fiecare dată rezultate diferite, deși foarte apropiate între ele. Aceste rezultate sînt determinate, în afară de factori *principali* — expresie a condițiilor *esențiale* în care are loc experimentul măsurătorii repetate a greutății — și de numeroși factori neesențiali, *întîmplători* (poziția corpului pe talerul cîntarului, vibrațiile întîmplătoare ale aparatului, variațiile fiziologice, în sensibilitatea simțurilor executantului măsurătorii etc.). Nu este exclus ca unul sau mai mulți din acești factori să apară doar în una sau cîteva operații de măsurare, dispărînd în celelalte.

Acțiunea factorilor întîmplători generează fenomene cunoscute sub denumirea de fenomene întîmplătoare. Prezența lor se înregistrează și în domeniul vieții social-economice. Datele cantitative ce caracterizează diferite fenomene din acest domeniu sînt rezultatul fie al unei singure cauze, fie al unei mulțimi de cauze. În particular, numărul cauzelor care generează un fenomen economic oarecare poate fi extrem de mare. Or, teoria probabilităților vizează, în principal, tocmai studiul informației cantitative *multicauzale*, obiectivul ei final fiind acela de a releva, din ansamblul aparent haotic al informației studiate, unele *tendințe* fundamentale, *legități* de bază.

Măsurînd pe baza unor modele abstracte conexiunea dialectică dintre necesitate și întîmplare, teoria probabilităților pune în evidență existența reală a unor *legități specifice* care guvernează ansambluri de fenomene întîmplătoare de masă, din natură și societate, denumite *legi probabilistice* (stohastice) sau legi de probabilitate.

În accepția filozofiei materialist-dialectice, există o conexiune obiectivă între necesitate și întîmplare, concretizată în trecerea uneia în cealaltă, în procesul dezvoltării materiei; mai mult, fenomenele întîmplătoare sînt o formă de manifestare a necesității și chiar o completare a ei. Așadar, fundamentul ontologic al legilor de probabilitate— legi cu caracter *obiectiv*, *necesar* și *general* — îl constituie dialectica necesității și întîmplării.

Descoperirea legilor de probabilitate este rezultatul unui proces îndelungat al activității științifice de abstractizare în domeniul teoriei probabilităților. Caracterul obiectiv al unei legi de probabilitate este relevat de faptul că fenomenul întîmplător i se supune, *independent* de voința noastră. Se spune că o lege de probabilitate este generală pentru că se manifestă în *toate* fenomenele și procesele cu caracter întîmplător.

Scanned with OKEN Scanner

În ceea ce privește caracterul de *necesitate* al legilor de probabilitate, el capătă un sens anumit în contexul teoriei probabilităților. Este știut că legile deterministe formulate de diferite științe sînt necesare în sensul că la fiecare realizare a unui ansamblu de condiții A are loc întotdeauna evenimentul B; legăturile legice dintre fenomene acționează deci — în anumite condiții — cu forța necesității. De pildă, legea căderii corpurilor descoperită de Galileu acționează cu forța necesității: toate corpurile și întotdeauna cad pe pămînt cu o accelerație de 9,8 m/s.

În domeniul fenomenelor întîmplătoare, la îndeplinirea unui ansamblu de condiții S, nu rezultă cu necesitate evenimentul A, însă rezultă în mod *necesar* — la reiterarea experimentului — *o stabilizare* a frecvențelor de apariție a fenomenului întîmplător studiat în jurul unui anumit număr. Acest număr reprezintă elementul stabil — indicator caracteristic obiectiv al fenomenului studiat —, măsură a *probabilității* lui de apariție.

În continuare, vom nota prin p sau $P(A)$, $P(B)$ probabilitatea ca la îndeplinirea sistemului de condiții S să se realizeze evenimentul întîmplător A, B etc.

Să ne referim, de pildă, la experimentul clasic constînd în aruncarea repetată, de *un mare număr* de ori, a monedei. În acest caz se poate observa că frecvența de apariție a unei fețe pe care o avem în vedere se stabilizează, convergînd, oscilatoriu, către un număr bine determinat, și anume 1/2. Asemenea proprietăți de stabilizare a frecvențelor pot fi observate și la alte experimente repetate de un mare număr de ori (stabilizarea frecvenței de apariție a unei fețe a zarului, a unei bile de o anumită culoare extrasă din urnă etc.), relevînd existența concretă a unor legități aparte, cunoscute sub denumirea generică de *legități statistice*. Aceste legități exprimă ceea ce este legic necesar pentru întreaga mulțime de fenomene întîmplătoare, considerată în ansamblul ei. Legile de probabilitate reflectă pe plan abstract tocmai aceste legități statistice reale, obiectiv existente, care guvernează fenomenele întîmplătoare de masă din natură și societate. Ele nu ne oferă posibilitatea de a prevedea evoluția unui fenomen întîmplător oarecare, luat izolat, ci fac posibilă previziunea unui rezultat global, *mediu,* pe ansamblul omogen al fenomenelor întîmplătoare de masă supuse cercetării. Legile de probabilitate reprezintă astfel un instrument de investigație de interes major pentru dezvoltarea, în general, a cercetărilor fundamentale și aplicative. [1] În unele domenii ale fizicii și mecanicii de pildă, mai ales la nivelul cuantic al materiei, modelele teoriei probabilităților se dovedesc a fi instrumente de investigație unice pentru descoperirea legilor care guvernează materia respectivă.

Experiment. Probă. Eveniment. Teoria probabilităților folosește o serie de concepte fundamentale cum sînt: experimentul, proba și evenimentul.

Realizarea practică a unui complex de condiții S, corespunzător unui criteriu dat de cercetare a diferitelor populații statistice se numește *experiment*, iar alegerea la întîmplare a unui element al populației pentru a i se identifica cantitativ sau calitativ criteriul de apartenență la populația considerată se numește *probă.*

Rezultatul unei probe este generalizat, teoretic, prin conceptul de *eveniment.*

[1] O expunere mai amplă privind obiectul, metodele și fundamentarea teoriei probabilităților poate fi găsită în [63], [83], [84] și [85].

Scanned with OKEN Scanner

Ne propunem, de pildă, să cercetăm calitatea unor piese de același tip, produse de o întreprindere industrială. Organizăm în acest scop producția în cicluri de fabricație repetate și în timpul procesului de fabricație extragem la întîmplare cîte o piesă din serii de cîte o sută. Fiecare piesă extrasă este supusă controlului de calitate, stabilindu-se în ce măsură caracteristicile ei tehnice corespund unor anumite standarde. În funcție de constatarea făcută, le declarăm fie piese de bună calitate, fie piese rebut. În acest exemplu identificăm imediat că populația statistică o constituie piesele; ea se constituie ca atare după criteriul unei anumite calități — caracteristici tehnice identice — care o omogenizează. De asemenea, identificăm conținutul material al conceptelor de experiment, probă și eveniment, constînd, respectiv, din realizarea practică a procesului de fabricație în cicluri identice (experiment), alegerea la întîmplare a unei piese dintr-o serie de o sută (probă) și rezultatul constatat — piesă de calitate sau piesă rebut (eveniment).

Evenimentele se clasifică în trei mari categorii: evenimente sigure, evenimente imposibile și evenimente întîmplătoare.

Prin eveniment *sigur* se înțelege acel eveniment care se produce în mod obligatoriu la efectuarea unei probe. Este un eveniment sigur, de pildă, extragerea unei bile albe dintr-o urnă conținînd numai bile albe; tot un eveniment sigur este extragerea dintr-o urnă cu bile de culoare albă și neagră, fie a unei bile albe, fie a unei bile negre. În exemplul considerat mai înainte, referitor la controlul statistic al calității pieselor, extragerea într-o probă fie a unei piese standard, fie a unei piese rebut constituie, de asemenea, un eveniment sigur.

Evenimentul contrar evenimentului sigur este un eveniment *imposibil*, adică un eveniment care în mod obligatoriu *nu* va avea loc la efectuarea unei singure probe. Extragerea unei bile negre dintr-o urnă ce conține numai bile albe este, evident, un eveniment imposibil.

Se numește un eveniment *întîmplător* (aleator) acela care poate fie să se producă, fie să nu se producă la efectuarea unei probe. Să considerăm experimentul clasic al aruncării unui zar „corect", adică a unui zar ideal, perfect simetric și confecționat dintr-un material omogen. Dacă aruncăm o singură dată un asemenea zar este evident că *oricare* din cele șase fețe ale sale poate să apară sau să nu apară. Nu putem anticipa, de pildă — dorind un răspuns *unic*, bine *determinat* — dacă după efectuarea unei singure probe va apărea fața pe care sînt imprimate șase puncte. Apariția acestei fețe — ca și a celorlalte cinci — constituie, prin definiție, un eveniment întîmplător.

Așadar, referitor la problema de a ști dacă evenimentul întîmplător se va produce sau nu, într-o singură probă, nu putem da un răspuns anticipat cert, bine determinat. Cu totul altfel se prezintă însă problema dacă avem în vedere evenimente întîmplătoare ce pot fi observate, în experimentări repetate. Asemenea evenimente de masă se supun — după cum am văzut — unor legități specifice.

Vom ilustra această idee și printr-un exemplu din activitatea economică.

Doi muncitori lucrează în aceeași secție dintr-o întreprindere industrială, la două strunguri de același tip și în aceeași stare de funcționare. Ambii execută aceleași operații de prelucrare a unor piese metalice, productivitatea muncii lor fiind apreciată după numărul de piese prelucrate într-o unitate de timp. Unul dintre ei, fruntaș în producție, cu o înaltă calificare tehnică și o bogată experiență în producție, obține sistematic o productivitate medie

Scanned with OKEN Scanner

a muncii mai înaltă decît celălalt muncitor, care are aceeaşi calificare, însă nu are încă experienţa şi îndemînarea primului. La o anumită oră a zilei de muncă, aleasă la întîmplare, se înregistrează producţia celor doi muncitori şi se constată că muncitorul fruntaş în producţie a prelucrat patru piese, iar celălalt şase piese. Cum se poate explica un asemenea rezultat? Este evident că cel care cunoaşte exact condiţiile în care lucrează cei doi muncitori, gradul lor de îndemînare şi însuşirile lor personale, va răspunde fără întîrziere că acest rezultat este „întîmplător".

Ne întrebăm, totuşi, cum s-ar putea scoate în evidenţă faptul că, în condiţiile date, primul muncitor are o productivitate orară medie a muncii mai înaltă? Şi la această întrebare răspunsul îl putem da simplu, intuitiv: vom face un *număr suficient de mare* de asemenea încercări, înregistrînd, de pildă, producţia executată de fiecare muncitor în opt ore succesive. Sporind astfel numărul observaţiilor, vom constata, la sfîrşit, că productivitatea orară a fruntaşului în producţie este semnificativ superioară şi putem spune că tocmai acest rezultat se impune ca fiind *legic*, *necesar*.

Acest exemplu ne sugerează intuitiv şi *condiţia* care trebuie îndeplinită — observarea de masă — pentru ca legitatea ascunsă, care guvernează fenomenele întîmplătoare, să fie pusă în evidenţă.

În teoria probabilităţilor s-au constituit cîteva concepte specifice, definind unele categorii de evenimente întîmplătoare.

a) Evenimentele întîmplătoare pot fi *incompatibile* şi *compatibile*.

Două evenimente se numesc *incompatibile*, dacă nu se pot realiza simultan; realizarea unuia exclude realizarea celuilalt. De pildă, extragerea unei piese dintr-un lot care conţine piese standard şi piese rebut ne conduce fie la evenimentul A — ca piesa să fie standard, fie la evenimentul B — ca piesa să fie rebut.

Evenimentele *compatibile* sînt acele evenimente care pot avea loc în acelaşi timp. Se poate considera, astfel, că apariţia feţei 6 la aruncarea zarului — evenimentul A — şi apariţia unei feţe pare — evenimentul B — constituie două evenimente compatibile. Ne putem referi, de pildă, şi la două populaţii de piese, conţinînd, fiecare, piese de calităţi diferite. Evenimentele de a extrage din prima populaţie o piesă de prima calitate, iar din a doua, o piesă de calitate inferioară, sînt evenimente compatibile.

b) *Evenimente dependente* şi *evenimente independente*. Evenimentele întîmplătoare se spune că sînt *dependente* între ele dacă probabilitatea de apariţie a unuia din ele se schimbă în funcţie de apariţia sau neapariţia celorlalte evenimente. Dacă realizarea unuia din evenimente nu influenţează probabilitatea realizării altor evenimente se spune că avem evenimente *independente*. Exemple de evenimente dependente şi independente vor fi date mai jos, la prezentarea unor modele clasice de calcul al probabilităţilor.

Operaţii cu evenimente. Vom considera, în acest context, cîteva elemente strict necesare din teoria algebrei Boole, ca şi unele notaţii ce se folosesc în teoria mulţimilor, pentru a fi în măsură să interpretăm evenimentele întîmplătoare ca elemente ale unor mulţimi şi, totodată, să abordăm, în continuare, unele probleme privind construcţia axiomatică a teoriei probabilităţilor.

În primul rînd, cîteva explicaţii introductive pe marginea unor reprezentări geometrice.

Scanned with OKEN Scanner

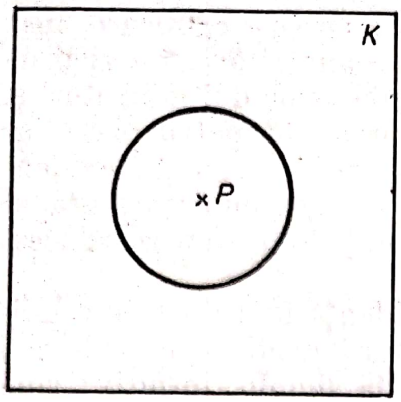


Fig. 3.1

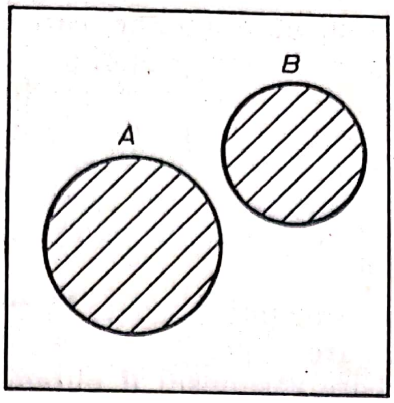


Fig. 3.2

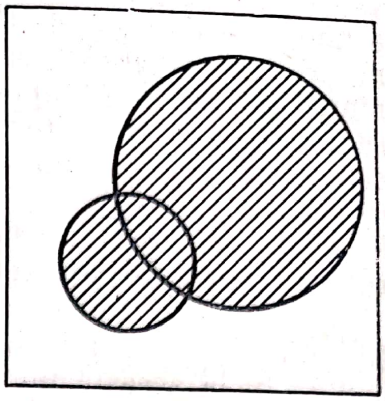


Fig. 3.3

Să considerăm astfel pătratul K din fig. 3.1 de mai sus și să notăm cu E evenimentul sigur că un punct, luat la întîmplare, aparține acestui pătrat. Un punct exterior pătratului, de pildă, va fi evenimentul imposibil Φ deoarece el nu-i poate aparține.

Faptul că un punct interior pătratului este interior unui cerc dat A, cuprins în K, constituie evenimentul A. Deoarece realizarea evenimentului A, adică punctul P, se găsește în cercul A, atrage simultan și realizarea evenimentului E; vom spune că evenimentul A implică evenimentul E. Această relație de implicație se scrie astfel.

$$A \subset E:$$

Relația „$\subset$" se numește *incluziune.* În general dacă avem relația de incluziune:

$$A \subset B \text{ sau } B \supset A$$

se spune că A este un *subelement* al lui B, sau că A este *conținut* în B, sau că B *conține* pe A; se mai spune, de asemenea, că A este mai *mic* decît B sau că B este un *majorant* al lui A, iar A un *minorant* al lui B.

Dacă facem o probă experimentală, alegînd la întîmplare un punct al lui K, el se va găsi sau *în* A sau în exteriorul lui A. Notînd acest din urmă eveniment cu A^* vom spune că evenimentele A și A^* sînt evenimente *opuse* sau *complementare.* Într-o probă, dacă se realizează A, nu se realizează A^* și reciproc.

Se observă că evenimentele opuse sînt *incompatibile.* Evenimentul opus lui A se notează simbolic cu $E-A$, adică:

$$A^* = E - A$$

sau și $\complement A$ înțelegînd prin aceasta, mulțimea punctelor pătratului, exterioare cercului A.

Realizarea a cel puțin unuia din evenimente: sau A, sau B, o vom numi *reuniunea* evenimentelor A și B. Obținem astfel un nou eveniment S, care se notează astfel:

$$S = A \cup B$$

care se citește „S egal A reunit cu B".

În figurile 3.2. și 3.3. sînt reprezentate cercurile A și B. Evenimentul reunit A sau B este acela ca un punct din interiorul pătratului, luat la întîmplare, să aparțină interiorului hașurat al celor două cercuri.

Scanned with OKEN Scanner

Deci evenimentul ca un punct al pătratului să aparţină regiunii haşurate constituie reuniunea $A \cup B$.

Din fig. 3.2 se vede că evenimentele A şi B sînt incompatibile, deoarece realizarea, într-o singură probă, a evenimentului A exclude realizarea evenimentului B şi invers. Evenimentele A şi B din fig. 3.3 sînt compatibile, căci alegerea unui punct *comun* celor două cercuri atrage după sine realizarea atît a evenimentului A, cît şi a evenimentului B.

În legătură cu operaţia de reuniune a evenimentelor sînt de făcut două observaţii importante:

1°. Dacă $A \subset B$, atunci reuniunea evenimentului A şi B este egală cu evenimentul B: $A \cup B = B$.

2°. Oricare ar fi evenimentul A, avem relaţiile:

$$A \cup A = A$$

$$A \cup \Phi = A$$

$$A \cup E = E$$

$$E \cup \Phi = E.$$

Realizarea *simultană* a două evenimente A şi B înseamnă un *nou* eveniment P pe care-l vom denumi *intersecţia* evenimentelor A şi B. Intersecţia P o vom nota astfel:

$$P = A \cap B$$

citind-o: evenimentul A intersectat cu evenimentul B. Din punct de vedere geometric, intersecţia $A \cap B$ este reprezentată prin regiunea haşurată *comună* celor două cercuri din fig. 3.3.

Prin introducerea operaţiilor de reuniune şi intersecţie, unele concepte din teoria probabilităţilor pot fi formulate în mod mai precis. Astfel, pentru evenimente complementare avem următoarea definiţie:

Evenimentele A şi $\complement A$ se numesc *complementare* dacă satisfac relaţiile:

$$A \cup \complement A = E \quad \text{şi} \quad A \cap \complement A = \Phi.$$

Evenimentele A şi B sînt *incompatibile* între ele dacă nu se pot realiza simultan, adică

$$A \cap B = \Phi.$$

În caz contrar, dacă avem relaţia

$$A \cap B \neq \Phi$$

evenimentele sînt *compatibile.*

Frecvenţă relativă şi probabilitate. Vom defini, în primul rînd, conceptul clasic de probabilitate pornind de la un exemplu:

La o întreprindere industrială oarecare s-a constatat că, în medie, 2% din piesele produse de o maşină automată sînt necorespunzătoare. Aceasta înseamnă că la fiecare partidă de produse nesortate piesele rebutate vor fi în proporţie de aproximativ 2%. Dacă partidele sînt formate, de exemplu, din 1 000 de piese, la unele dintre ele numărul rebuturilor va fi sub 2% (16, 17,

Scanned with OKEN Scanner

..., piese), la altele peste 2% (22, 23, ..., piese) dar, în medie, acest număr va fi apropiat de 20.

Să presupunem că procesul de fabricaţie are loc în aceleaşi condiţii de producţie. În acest caz, operaţia de masă constă în fabricarea în serie a produselor, conducînd la constituirea unei populaţii omogene. Procentul pe care îl poate forma numărul de apariţii al unuia sau al altuia dintre evenimentele care ne interesează (produse necorespunzătoare) va fi, în condiţii de producţie identice, aproximativ acelaşi, abătîndu-se de la o anumită *valoare medie*, relativ stabilă, cu foarte puţin. Se spune că această valoare medie este indicatorul caracteristic al operaţiei de masă; ea indică numărul produselor rebut, avînd probabilitatea, aproximativ constantă, p.

Este foarte importantă cunoaşterea acestui indicator în diferitele domenii de activitate. El face posibilă aprecierea fenomenelor de masă întîmplătoare şi chiar previziunea evoluţiei lor viitoare, în măsura în care condiţiile iniţiale ale experimentului rămîn aceleaşi.

În exemplul de mai înainte, în care la 1 000 de piese, produse de o maşină automată, 20 de piese sînt în medie rebut, se spune că probabilitatea p de a produce rebuturi este, pentru maşina dată, $\frac{20}{1\,000} = 0{,}02$.

Prin urmare, într-o operaţie de masă care are loc în condiţii identice, un eveniment A se produce, în medie, de m ori, adică la m din N unităţi elementare ale populaţiei studiate. Probabilitatea evenimentului A este:

$$P(A) = \frac{m}{N}. \tag{3.1}$$

În această relaţie, N reprezintă numărul cazurilor egal posibile, pe cînd m reprezintă numărul cazurilor favorabile producerii evenimentului A; ea sintetizează definiţia clasică a conceptului matematic de probabilitate:

Se numeşte probabilitatea unui eveniment şi se notează cu $P(A)$ raportul dintre numărul m de rezultate favorabile producerii lui A şi numărul total N de rezultate posibile ale experimentului, cu condiţia ca toate rezultatele să fie egal posibile.

Pe baza acestei definiţii se vede imediat că probabilitatea de apariţie, la o singură aruncare, a uneia din feţele unui zar omogen şi perfect construit este $\frac{1}{6}$ sau probabilitatea de apariţie a uneia din feţele monedei este $\frac{1}{2}$ etc. Deoarece $m \leqslant N$ rezultă că probabilitatea oricărui eveniment întîmplător A satisface dubla inegalitate

$$0 \leqslant P(A) \leqslant 1. \tag{3.2}$$

Cu cît $P(A)$ este mai apropiat de 1, cu atît evenimentul A are loc mai des. Dacă $P(A) = 0$, evenimentul A sau nu are loc niciodată, sau are loc foarte rar, aşa că practic îl considerăm imposibil. Dacă $P(A) = 1$, evenimentul are loc totdeauna — deci este un eveniment sigur.

Din definiţia clasică a probabilităţii (3.1) rezultă deci următoarele proprietăţi:

1°. Probabilitatea evenimentului *sigur* este 1, întrucît în acest caz avem $m = N$;

2°. Probabilitatea evenimentului imposibil este 0, întrucît în acest caz $m = 0$;

3°. Probabilitatea unui eveniment întîmplător este cuprinsă între 0 și 1, întrucît în acest caz $0 < m < 1$.

În afară de conceptul de probabilitate există în teoria probabilităților un alt concept fundamental, și anume cel de *frecvență relativă*. Prin frecvența relativă a evenimentului A se înțelege raportul dintre numărul probelor m în care evenimentul A s-a produs experimental și numărul total N de probe efectuate.

Dintr-o îndelungată observație a fenomenelor și proceselor de masă s-a putut constata că dacă un experiment se repetă, în aceleași condiții, de un număr suficient de mare de ori, atunci frecvența relativă capătă o anumită stabilitate, oscilînd în jurul probabilității. Tocmai de aceea, drept măsură cantitativă de apreciere a posibilității obiective de a se produce evenimentul întîmplător A poate fi luată frecvența relativă f_A rezultată după un număr mare de N experimente, efectuate în aceleași condiții.

După cum se vede, noțiunea de probabilitate a unui eveniment este legată, chiar la originea formării ei, de o noțiune experimentală, practică — frecvența evenimentului —, rezultată din legile obiective ale fenomenelor reale de masă. Aceasta a condus la constatarea că evenimentele corespunzătoare diferitelor probe experimentale formează o anumită structură, cu numeroase proprietăți care pot fi formulate matematic. Matematicianul sovietic A. N. Kolmogorov a numit-o *cîmp de evenimente* și pe această bază a formulat cunoscutele lui axiome privind teoria probabilităților.

Definiția clasică a probabilității dată de raportul dintre numărul cazurilor favorabile și numărul cazurilor egal posibile (adică egal probabile) se bazează, esențial, pe discriminarea cazurilor egal posibile. Or, noțiunea de „egală probabilitate" care intervine în această definiție nu este nici ea definită. Sîntem deci într-un cerc vicios, dat fiind că definiția probabilității se reduce la definiția categoriei mai simple a „echiprobabilității", care nu poate fi definită din punct de vedere matematic, formal, ci doar ilustrată.

Această situație în care se evidențiază limitele gnoseologice ale definiției clasice a probabilității a dus la necesitatea axiomatizării ei.

La începutul deceniului al patrulea, A. N. Kolmogorov, legînd teoria clasică a probabilităților de teoria metrică a funcțiilor și de teoria mulțimilor, a pus bazele teoriei moderne a probabilităților [2]. Axiomele luate în considerare de Kolmogorov nu sînt niște elaborări abstracte ale gîndirii, ele se bucură de proprietatea de a nu fi contradictorii și, totodată, nici una din ele nu este o consecință a alteia.

Conceptele de bază folosite în axiomatizarea lui Kolmogorov sînt evenimentele și probabilitățile respective.

Axiomele teoriei probabilităților sînt următoarele:

1° *Unui experiment îi corespunde întotdeauna un cîmp de evenimente.* De pildă, experimentul aruncării o singură dată a unui zar generează un cîmp de evenimente format din $2^6 = 64$ evenimente, elementare, din care 62 se constituie din *realizările posibile* ale uneia din cele șase fețe ale zarului (1),

[2] Trebuie să arătăm că și F.P. Cantelli s-a preocupat de axiomatizarea teoriei probabilităților într-o lucrare din 1932, *Una teoria astrata del calcolo delle probabilita*. Ea a fost însă desăvîrșită, cu un an mai tîrziu, de A.N. Kolmogorov.

(2), ..., (6), ale uneia din două fețe (1, 2), (1, 3), ..., (5, 6), ale uneia din trei fețe (1, 2, 3), (1, 2, 4),.... (4, 5, 6) ș.a.m.d. Se mai adaugă apoi evenimentul *sigur*, constînd în faptul că în mod sigur va apărea una din fețele zarului și evenimentul *imposibil*, constînd în faptul imposibil ca să nu apară nici una din aceste fețe.

În general, un cîmp *finit* de evenimente este caracterizat prin următoarele proprietăți:

a) Dacă notăm cu A_k $(1 \leqslant k \leqslant h)$ evenimente ale cîmpului, atunci $\bigcup_{k=1}^{h} A_k$ și $\bigcap_{k=1}^{h} A_k$ sînt și ele evenimente ale cîmpului.

b) Notînd A_k^*, evenimentul complementar lui A_k, atunci și A_k^* este un eveniment al cîmpului.

Pentru un cîmp infinit trebuie să se admită că și $\bigcap_{k=1}^{\infty} A_k$, $\bigcap_{k=1}^{\infty} A_k (k \geqslant 1)$ sînt evenimente ale cîmpului considerat.

2° *Fiecărui eveniment al cîmpului A îi corespunde un număr real, nenegativ, subunitar p(A), adică*

$$0 \leqslant p(A) \leqslant 1;$$

acest număr se numește probabilitatea evenimentului A.

3° *Probabilitatea evenimentului sigur este egală cu 1.*

$$P(E) = 1.$$

4° *Dacă evenimentele A, B, ... sînt incompatibile, atunci*

$$p(A + B + ...) = p\ (A) + p(B) + ...$$

Semnul „+" indică reuniunea $(\cup)$ și se citește „sau".

5° *Dacă evenimentele A, B, ... sînt independente, atunci*

$$p(A \cdot B ...) = p(A) \cdot p(B) ...$$

Semnul „·" indică intersecția $(\cap)$ și se citește „și".

Axiomatizarea teoriei probabilităților formulată de Kolmogorov a înlăturat lipsurile definiției clasice a conceptului de probabilitate.

Definiție: Probabilitatea este o măsură a posibilității obiective de realizare a unui eveniment întîmplător.

Formularea acestei definiții a deschis largi posibilități pentru abordarea a noi direcții de cercetare în domeniul complex al fenomenelor aleatoare. În ultimele patru decenii teoria probabilităților a înregistrat un considerabil salt calitativ în dezvoltarea ei.

Dacă avem un cîmp oarecare de s evenimente elementare sau complexe, în așa fel încît în fiecare probă se produce neapărat unul din ele, și numai unul, vom spune că ansamblul acestor evenimente este un *sistem complet de evenimente.*

În termenii axiomelor a 4-a și a 3-a, ale lui Kolmogorov, avem:

$$A_1 \cup A_2 \cup A_3 ... \cup A_s = E,$$

Scanned with OKEN Scanner

eveniment sigur, avînd probabilitatea:

$$P(A_1 \cup A_2 \ldots \cup A_s) = p(E)$$

sau

$$P(A_1) + P(A_2) + \ldots + P(A_s) = 1,$$

adică suma probabilităţilor unor evenimente care formează un sistem complet de evenimente este egală cu 1.

Se poate arăta, de pildă, că două evenimente contrare (complementare) A şi A^* formează un sistem complet de evenimente şi deci suma probabilităţilor lor este egală cu 1.

Desigur, în cazul evenimentelor contrare sînt valabile, după cum am mai văzut, relaţiile

$$A \cup A^* = E \text{ şi } A \cap A^* = \Phi.$$

Probabilitatea evenimentului A, calculată condiţionat de realizarea evenimentului B, se numeşte *probabilitate condiţionată* şi se notează $P(A/B)$. Probabilitatea condiţionată prezintă un interes deosebit în cercetarea şi caracterizarea dependenţei a două variabile aleatoare, componente ale unor repartiţii multidimensionale.

Probabilitatea evenimentului A, condiţionată de realizarea evenimentului B, este dată de relaţia:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \qquad (3.3)$$

$$p(B) \neq 0.$$

Principiul imposibilităţii practice a evenimentelor cu probabilitatea mică. În rezolvarea multor probleme practice întîlnim evenimente a căror probabilitate este foarte mică. Putem considera că un eveniment A cu probabilitate foarte mică (de pildă, $p < 0{,}01$) nu se produce într-o singură probă.

Trebuie să observăm, mai întîi, că sînt cazuri (nu ne mărginim, bineînţeles, la definiţia clasică a probabilităţii) în care chiar un eveniment cu probabilitatea zero nu este teoretic imposibil.

Să presupunem astfel, că, într-o urnă ideală, sînt scrise pe bilete separate toate numerele reale. Ne întrebăm care este probabilitatea ca extrăgînd, la întîmplare, un bilet să obţinem numărul 2?[3]. Dat fiind faptul că mulţimea numerelor reale este infinită, probabilitatea ca să obţinem numărul 2 este egală cu zero. De data aceasta, cu toate că probabilitatea este nulă, nu există certitudinea că nu vom putea obţine niciodată numărul 2.

Statisticienii obişnuiesc să considere un eveniment ca practic imposibil într-o singură probă dacă probabilitatea ca el să se producă este mai mică decît un anumit număr (de exemplu 0,03). Determinarea acestui *nivel de semnificaţie*, după denumirea pe care o are în statistica matematică, variază de la caz la caz, după importanţa şi consecinţele fenomenului considerat.

În domeniul producţiei industriale, putem afirma uneori că o serie de produse fabricate este acceptabilă dacă probabilitatea de a obţine un rebut

[3] Problema mai poate fi formulată în felul următor: pe axa numerelor reale se alege, la întîmplare, un număr; care este probabilitatea ca acest număr să fie 2?

Scanned with OKEN Scanner

este egală cu 1‰. Nu întotdeauna însă ne putem mulțumi cu asemenea limită. Necesitățile practicii impun uneori un nivel de semnificație egal cu 10^{-5} sau 10^{-6} și chiar mai mic, echivalînd, practic, cu imposibilitatea apariției evenimentului, care, în general, nu este de dorit să apară în domeniul respectiv de activitate.

Ceea ce s-a arătat relativ la evenimentele cu probabilitatea foarte mică, ne conduce la următoarea afirmație: dacă un eveniment întîmplător are probabilitatea foarte apropiată de 1, ($p > 0{,}99$), atunci putem considera că acest eveniment este practic sigur într-o singură probă.

3.2 VARIABILE ALEATOARE UNIDIMENSIONALE ȘI REPARTIȚIILE LOR

3.2.1 Noțiuni introductive. Definiții

În teoria probabilităților, conceptul de variabilă aleatoare [4] este fundamental.

În viața de toate-zilele întîlnim frecvent mărimi care iau valori ce se schimbă sub influența unor factori întîmplători. Așa sînt, de exemplu, numărul de zile dintr-un an în care cade ploaia într-o anumită regiune, numărul băieților din 100 nou născuți, masa unui bob de fasole luat dintr-o anumită recoltă, rezultatul obținut în urma măsurării unei mărimi fizice, viteza unei molecule de gaz dintr-un recipient, recolta medie la un hectar, mortalitatea unui grup de indivizi etc.

Mărimile care iau valori la întîmplare sînt legate *de anumite experimente*; valoarea pe care o ia mărimea respectivă este în funcție de rezultatul experimentului de care este legată. O astfel de mărime, definită printr-un număr, legată de un experiment aleator și care ia valori la întîmplare, în funcție de rezultatul experimentului, poartă denumirea de *variabilă aleatoare*.

Unele variabile aleatoare, numite *discrete*, pot lua un număr finit de valori, în timp ce altele, numite *continue*, pot lua orice valoare dintr-un interval (a, b).

Variabilele din prima categorie le vom numi *discrete*, iar variabilele din categoria a doua, — *continue*.

În afară de variabilele continue mai există și alte variabile aleatoare care iau o infinitate de valori. De pildă, în cazul repartiției Poisson, șirul valorilor posibile este infinit (0, 1, 2, ...).

Exemple de variabile aleatoare *discrete:*

— numărul de chemări pe care centrala telefonică automată le primește într-un minut (0, 1, 2, ...);

— numărul de nimeriri la țintă într-un experiment de trei trageri (0, 1, 2, 3,);

— numărul de copii în diferite familii din R.S.R. (0, 1, 2, ...).

[4] În practică, pentru o variabilă aleatoare oarecare se mai folosește și denumirea de variabilă întîmplătoare sau *caracteristică cantitativă* variabilă și, după caz, *caracteristică atribut*. În capitolul următor consacrat definirii legilor de repartiție pe baza datelor empirice vor fi folosite tocmai aceste noțiuni.

Scanned with OKEN Scanner

Variabilele aleatoare le simbolizăm prin litere mari, iar valorile lor posibile — prin litere mici. De pildă, dacă notăm cu X variabila aleatoare care reprezintă numărul de zile din cursul unei săptămîni, în care consumul de apă este normal, atunci această variabilă va putea lua următoarele șapte valori:

$X = x_1 = 0$ (nici o zi cu consum normal);

$X = x_2 = 1$ (o zi cu consum normal);

$X = x_3 = 2$ (2 zile cu consum normal);

.....

$X = x_7 = 6$ (6 zile cu consum normal).

Fiecare din aceste valori este posibilă dar nicidecum singură, ci are o anumită probabilitate pe care o vom nota cu $p_i (i = 1, 2, ..., 7)$; evenimentele $X = x_i$ $(i = 1, 2, ..., 7)$ sînt incompatibile.

În cazul general, cînd o variabilă aleatoare X poate lua un număr oarecare de valori $x_1, x_2, ..., x_n$, fiecare cu probabilitatea respectivă $p_1, p_2, ..., p_n$ este evident că apariția uneia din valori exclude pe celelalte.

Sîntem deci în prezența unui sistem complet de evenimente:

$$P(X = x_1) = p_1;\ P(X = x_2) = p_2;\ \ldots . P(X = x_n) = p_n$$

pentru care suma probabilităților tuturor valorilor posibile ale variabilei aleatoare X este egală cu 1;

$$\sum_{i-1}^{n} p_i = 1.$$

Dacă notăm evenimentele sistemului cu $A_1, A_2, ..., A_n$ putem defini astfel conceptul de variabilă aleatoare de tip discret.

O mărime X, care variază la întîmplare și poate lua, cu probabilitățile cunoscute p_i $(i = 1, 2, ..., n)$, una din valorile x_i $(i = 1, 2, ..., n)$, după cum se verifică unul din evenimentele A_i $(i = 1, 2, ..., n)$, formînd un sistem complet de evenimente, se numește variabilă aleatoare discretă.

În termeni mai simpli, exprimînd aceeași idee, putem spune:

Se numește variabilă aleatoare o mărime care, drept rezultat al unui experiment, poate lua o valoare oarecare, fără să se poată preciza dinainte care anume.

Exemple de variabile aleatoare *continue:* durata de ardere a unui bec; dacă durata minimă de garanție este de 500 ore, variabila aleatoare X poate lua orice valoare ce depășește 500; greutatea unui bob de grîu luat la întîmplare; eroarea de măsurare a unui corp oarecare pe cîntarul analitic; durata vieții unei persoane.

Este evident că în aceste exemple variabila aleatoare X are o mulțime infinită de valori, completînd integral un anumit interval și în această situație nu se poate vorbi în general de probabilitatea p_i a unei valori izolate x_i. Se poate afirma, în schimb, că există anumite domenii ale valorilor posibile ale variabilelor aleatoare continue X, care sînt luate cu o anumită probabilitate. După cum vom vedea mai jos, pentru caracterizarea variabilelor aleatoare continue este mai comod să nu folosim probabilitatea evenimentului $X = x$, ci probabilitatea evenimentului $X < x$.

Scanned with OKEN Scanner

3.2.2 Repartiția variabilei aleatoare discrete. Poligonul de repartiție

Faptul că suma probabilităților tuturor valorilor posibile ale unei variabile aleatoare este egală cu 1 ne sugerează în mod necesar ideea că această sumă se repartizează într-un mod oarecare între aceste valori. Din punct de vedere probabilistic, o variabilă aleatoare este complet caracterizată dacă se poate da o astfel de repartiție, care să indice cu precizie probabilitatea fiecăreia din valorile ei posibile.

Prin aceasta stabilim ceea ce vom denumi *lege de repartiție*[5] a variabilei aleatoare sau, pe scurt, *repartiție*.

Una din formele cele mai simple în care putem prezenta o lege de repartiție a probabilităților $p_1, p_2, \ldots, p_n$, corespunzătoare mulțimii de valori $x_1, x_2, \ldots x_n$ pe care le poate lua o variabilă aleatoare discretă X, este aceea a unui tabel pe care-l vom denumi *serie de repartiție:*

x_i	x_1	x_2	$\ldots x_i \ldots$	x_n
p_i	p_1	p_2	$\ldots p_i \ldots$	p_n

O altă formă de prezentare este cea grafică: pe axa absciselor sînt notate valorile posibile ale variabilei aleatoare, iar pe axa ordonatelor, probabilitățile corespunzătoare; pentru mai multă expresivitate, punctele obținute se unesc prin segmente de dreaptă, obținîndu-se graficul de forma celui din fig. 3.4, denumit *poligonul de repartiție.*

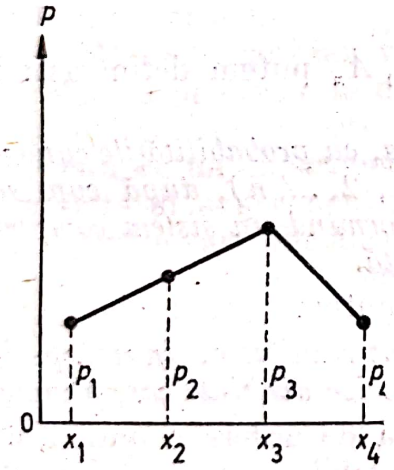


Fig. 3.4

Poligonul de repartiție, la fel ca și seria de repartiție, caracterizează, în întregime, o variabilă aleatoare discretă.

Aplicația 1. Să presupunem că se efectuează un experiment în următoarele condiții.

O mașină-unealtă produce, în serie, anumite piese. Din mulțimea acestor piese se extrage, la întîmplare, una și se consideră că are loc evenimentul A, dacă piesa este rebut și, evident, are loc evenimentul A^*, (non A) dacă piesa

[5] Unii cercetători folosesc și denumirea de lege de probabilitate, sinonimă cu lege de repartiție. După părerea noastră, impropriu se folosește denumirea de lege de distribuție.

este corespunzătoare prescripțiilor de calitate. Evenimentul A are probabilitate 0,1, iar evenimentul A^* probabilitatea 0,9. Să asociem evenimentului A o variabilă aleatoare discretă X, reprezentînd numărul de realizări ale evenimentului A în experimentul efectuat. Această variabilă este bivalentă. Dacă evenimentul A are loc, variabila aleatoare ia valoarea 1, iar în cazul cînd acest eveniment nu are loc, variabila aleatoare ia valoarea 0.

Se cere să construim seria de repartiție și poligonul de repartiție al variabilei aleatoare X.

Soluție. Variabila aleatoare X fiind bivalentă, seria de repartiție are următoarea formă:

x_i	0	1
p_i	0,9	0,1

Poligonul de repartiție este ilustrat în fig. 3.5.

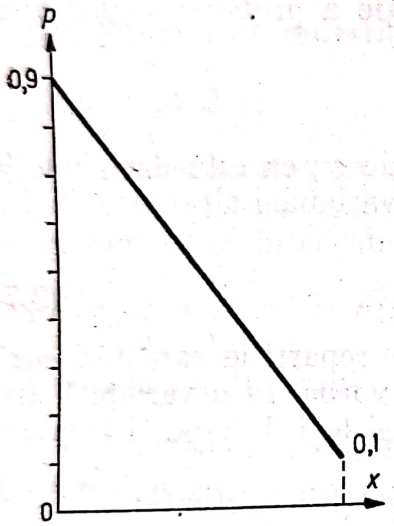


Fig. 3.5

În general, dacă ținem seama că $p_i = P\ (X = x_i)$ pentru oricare i = 1, 2, ..., n, atunci se poate spune că legea de repartiție a unei variabile aleatoare X exprimă probabilitatea ca variabila aleatoare să ia o anumită valoare x; probabilitatea este deci o funcție de valoare x:

$$P(X = x) = f(x) \quad \text{pentru} \quad x = x_1, x_2, \ldots, x_n \tag{3.4}$$

Funcția $f(x)$ fiind o probabilitate, este prin definiție nenegativă:

$$f(x) \geqslant 0. \tag{3.5}$$

Din faptul că după cum se realizează unul din evenimentele $A_1, A_2, \ldots, A_n$, formînd un sistem complet de evenimente, variabila aleatoare X, asociată lor, ia una din valorile $x_1, x_2, \ldots, x_n$, rezultă egalitatea:

$$f(x_1) + f(x_2) + \ldots + f(x_n) = \sum_{i=1}^{n} f(x_i) = 1; \tag{3.6}$$

condițiile (3.4) și (3.5) sînt îndeplinite întotdeauna.

Scanned with OKEN Scanner

3.2.3 Repartiția variabilei aleatoare continue. Funcție de repartiție

Dacă pentru variabile aleatoare discrete legea de repartiție constituie un instrument de caracterizare completă, nu același lucru se poate spune despre variabilele aleatoare continue. Este evident că nu putem caracteriza o asemenea variabilă printr-o lege de repartiție, prezentată sub forma unei serii sau a poligonului de repartiție, în sensul pe care l-am întîlnit la variabile aleatoare discrete. Avînd un număr *infinit* de valori, nu putem pune fiecare valoare în parte în evidență, cu probabilitatea ei respectivă. Vom vedea, dealtfel, mai tîrziu, că este posibil ca fiecărei valori izolate a unei variabile aleatoare continue să nu-i putem asocia o probabilitate diferită de zero.

Am remarcat mai sus că, totuși, unui anumit domeniu de variație a valorilor posibile ale variabilei aleatoare continue îi corespunde o probabilitate finită. Aceasta ne sugerează ideea că și în cazul unei variabile aleatoare continue există o anumită repartiție a probabilităților.

Pentru a caracteriza cantitativ această repartiție a probabilităților vom considera probabilitatea

$$P(X < x).$$

Probabilitatea evenimentului $(X < x)$ depinde de x; ea este deci funcție de x și de aceea o vom numi *funcție de repartiție* a variabilei aleatoare X și o vom nota $F(x)$:

$$F(x) = P(X < x). \tag{3.7}$$

Din punct de vedere probabilistic, funcția de repartiție caracterizează *complet* o variabilă aleatoare, indiferent dacă este vorba de o variabilă discretă sau continuă. Ea este deci încă o formă a unei legi de repartiție și are un caracter general.

Funcția de repartiție se bucură de unele proprietăți importante:

1°. Funcția de repartiție este o funcție nedescrescătoare; dacă $x_2 > x_1$, atunci $F(x_2) \geqslant F(x_1)$.

Acest lucru rezultă din relația:

$$P(X < x_2) = P(X < x_1) + P(x_1 \leqslant X < x_2), \quad x_2 > x_1 \tag{3.8}$$

întrucît reuniunea evenimentelor incompatibile $X < x_1$ și $x_1 \leqslant X < x_2$ ne dă evenimentul $X < x_2$.

Din această egalitate deducem:

$$P(X < x_2) - P(X < x_1) = P(x_1 \leqslant X < x_2) \geqslant 0$$

deoarece $P(x_1 \leqslant X < x_2)$ este mai mare sau egală cu zero, fiind o probabilitate. Cum

$$P(X < x_2) = F(x_2), \quad P(x < x_1) = F(x_1),$$

deducem:

$$F(x_2) - F(x_1) \geqslant 0$$

sau $F(x_2) \geqslant F(x_1)$ pentru $x_2 > x_1$.

Scanned with OKEN Scanner

2°. Pentru cea mai mică valoare posibilă a variabilei aleatoare $(-\infty)$ funcția de repartiție este egală cu zero.

$$F(-\infty) = 0. \tag{3.9}$$

În mod intuitiv, producerea evenimentului $X < -\infty$ este imposibilă și deci:

$$F(-\infty) = P(X < -\infty) = 0.$$

Se poate demonstra că într-adevăr relația de mai sus e adevărată.

3°. Pentru cea mai mare valoare posibilă a variabilei aleatoare $(+\infty)$ funcția de repartiție este egală cu 1.

Într-adevăr, constituie un eveniment sigur acela că variabila aleatoare va lua una din valori care sînt mai mici decît cea mai mare dintre ele. Rezultă, după cum se poate dovedi riguros, că:

$$F(+\infty) = P(X < +\infty) = 1. \tag{3.10}$$

4°. Funcția de repartiție fiind o probabilitate satisface dubla inegalitate:

$$0 \leqslant F(x) \leqslant 1. \tag{3.11}$$

5°. Probabilitatea ca variabila aleatoare X să fie cuprinsă între x_1 și x_2 $(x_1 \leqslant X < x_2)$ este egală cu diferența dintre valorile funcției de repartiție la extremitățile intervalului considerat:

$$P(x_1 \leqslant X < x_2) = F(x_2) - F(x_1). \tag{3.12}$$

Această relație rezultă din formula (3.8).

Proprietățile funcției de repartiție pot fi ușor puse în evidență cu ajutorul interpretării geometrice.

Aplicația 2. Vom considera, de pildă, că variabila noastră aleatoare X este un punct aleator de pe axa $0x$ din fig. 3.6 și, drept rezultat al unui eveniment,

Fig. 3.6

poate ocupa un loc sau altul. În acest caz, funcția de repartiție $F(x)$ este probabilitatea ca punctul aleator X să cadă, după efectuarea experimentului, la stînga punctului x.

Dacă punctul x îl vom deplasa spre stînga pe axa absciselor este clar că, atunci cînd x va ajunge la limită, X nu va mai putea lua o valoare în stînga lui x, evenimentul fiind imposibil (proprietatea 2°).

În mod analog, dacă punctul x îl deplasăm spre dreapta, atunci la un moment dat el va ajunge la cea mai mare valoare $(+\infty)$ și atunci ne convingem ușor că $F(+\infty) = 1$, dat fiind că evenimentul $X < +\infty$ este un eveniment sigur (proprietatea 3°). Se verifică că funcția de repartiție a variabilei X este

Scanned with OKEN Scanner

nedescrescătoare (proprietatea 1°), iar domeniul ei de variație este cuprins între 0 și 1 (proprietatea 4°).

Funcția de repartiție a unei variabile discrete, reprezentată grafic, apare ca o funcție în trepte.

În fiecare din punctele ei de pe axa absciselor ea admite salturi ca în fig. 3.7 Suma tuturor salturilor este egală cu 1.

Aplicația 3. Dintr-o urnă cu 7 bile albe și 3 negre se fac 4 extracții succesive, înapoindu-se bila în urnă după fiecare extracție. Să se construiască funcția de repartiție a variabilei aleatoare X, reprezentînd numărul de apariții ale bilei negre.

Soluție: Folosind modelul cunoscut privind calculul probabilităților în cazul unei urne de tip Bernoulli cu două stări — bila revenind în urnă după fiecare extracție — variabila aleatoare X capătă forma următoarei serii de repartiție:

x_i	0	1	2	3	4
p_i	0,2401	0,4116	0,2646	0,0756	0,0081

Cunoscînd seria de repartiție putem construi acum și funcția de repartiție:

dacă $x \leqslant 0$ $F(x) = 0$;
dacă $0 < x \leqslant 1$ $F(x) = 0{,}2401$;
dacă $1 < x \leqslant 2$ $F(x) = 0{,}6517$;
dacă $2 < x \leqslant 3$ $F(x) = 0{,}9163$;
dacă $3 < x \leqslant 4$ $F(x) = 0{,}9919$;
dacă $x > 4$ $F(x) = 1{,}0000$.

Graficul acestei funcții de repartiție se vede în fig. 3.8 de mai jos.

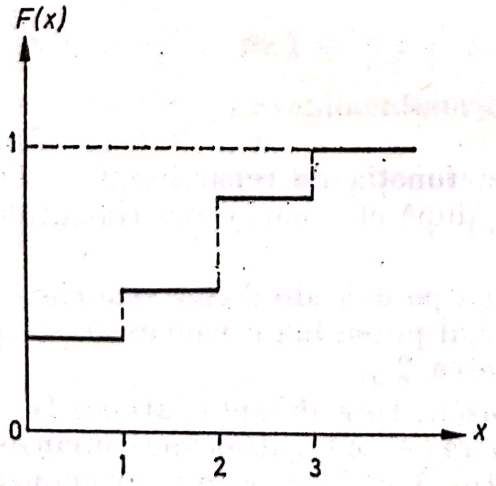


Fig. 3.7

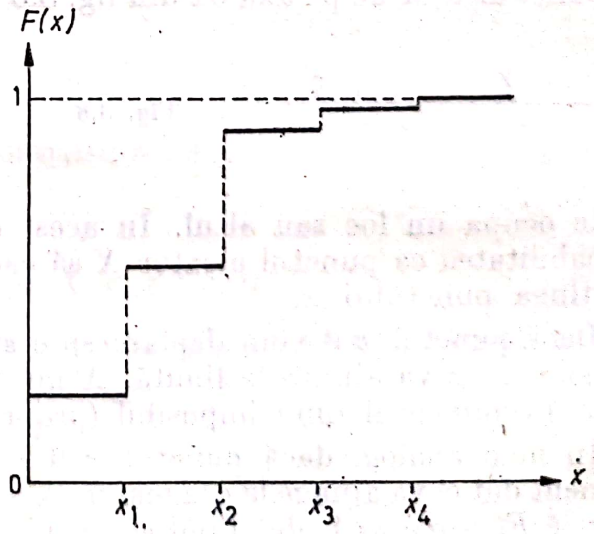


Fig. 3.8

Scanned with OKEN Scanner

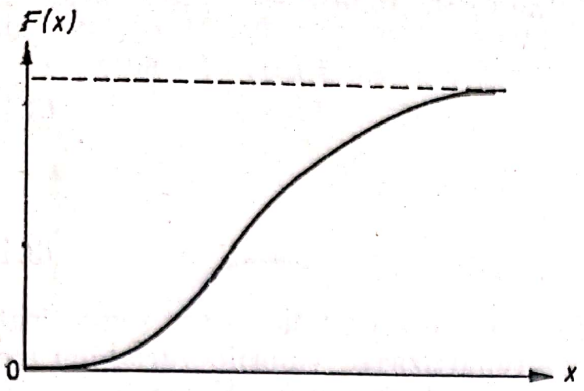


Fig. 3.9

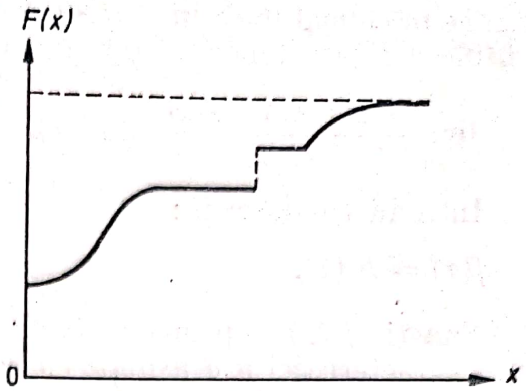


Fig. 3.10

În cazul variabilelor aleatoare de tip continuu, graficul funcției de repartiție este ilustrat ca în fig. 3.9, în care funcția are drept asimptote dreptele $F(x) = 0$ și $F(x) = 1$. În această situație, funcția de repartiție $F(x)$ este continuă și derivabilă.

Există și variabile aleatoare pentru care funcția de repartiție se reprezintă grafic ca în fig. 3.10.

În punctul x_0, $F(x)$ admite o discontinuitate. Variabilele aleatoare care admit o astfel de funcție de repartiție nu mai sînt de tip continuu.

Am anticipat mai sus, afirmînd că probabilitatea ca o variabilă aleatoare continuă să ia o anumită valoare individuală nu poate fi diferită de zero. Într-adevăr, ne putem convinge de aceasta folosindu-ne de relația:

$$F(\alpha \leqslant X < \beta) = F(\beta) - F(\alpha).$$

Făcînd pe β să tindă către α vom avea:

$$P(X = \alpha) = \lim_{\beta \to \alpha} P(\alpha \leqslant X < \beta) = \lim_{\beta \to \alpha} [F(\alpha) - F(\beta)] = 0.$$

Această relație se explică prin faptul că:

$$\lim_{\beta \to \alpha} F(\beta) = F(\alpha),$$

funcția de repartiție fiind continuă și derivabilă în cazul unei repartiții continue.

În cazul variabilelor aleatoare de tip continuu are sens să determinăm numai probabilitatea ca variabila aleatoare să fie cuprinsă într-un interval dat.

3.2.4 Densitate de repartiție

Să considerăm o variabilă aleatoare X, cu funcția de repartiție derivabilă $F(x)$. Vom calcula probabilitatea ca această variabilă să fie cuprinsă în intervalul: $(x,\ x + \Delta x)$, adică:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Raportînd această probabilitate la lungimea intervalului vom obține o probabilitate medie, pe unitatea de lungime a intervalului; dacă vom micșora

Scanned with OKEN Scanner

din ce în ce mai mult intervalul Δx, făcîndu-l să tindă către 0, la limită, vom obține derivata funcției de repartiție:

$$\lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x). \tag{3.13}$$

Introducem notația:

$$f(x) = F'(x). \tag{3.14}$$

Funcția $f(x)$ — prima derivată a funcției de repartiție — are semnificația unei caracteristici a *densității* cu care se repartizează valorile variabilei aleatoare în punctul dat; ea se numește *densitate de repartiție.* Poate fi folosit și termenul de *densitate de probabilitate.* Curba care ilustrează grafic densitatea de repartiție se numește *curba de repartiție.*

Ca și funcția de repartiție, densitatea de repartiție este una din formele legii de repartiție fără să fie generală: ea există numai pentru variabile aleatoare continue.

Din calculul diferențial știm că dacă intervalul elementar dx este suficient de mic, are loc egalitatea aproximativă:

$$F(x + \Delta x) - F(x) \approx f(x)\, dx.$$

În termenii teoriei probabilităților, vom spune că probabilitatea ca variabila aleatoare X să ia o valoare care să aparțină intervalului $(x, x + \Delta x)$ este aproximativ egală cu produsul densității de repartiție în punctul x cu lungimea dx a intervalului elementar.

Mai sus, prin relația (3.14), am arătat că densitatea de repartiție poate fi exprimată prin funcția de repartiție. Ne putem pune și o problemă inversă, aceea de a exprima funcția de repartiție prin densitatea de repartiție.

Din definiția funcției de repartiție rezultă că

$$F(x) = \int_{-\infty}^{x} f(x)\, dx + C,$$

unde C este o constantă. Se poate arăta că $C = 0$, făcînd pe x să tindă către $-\infty$; ținînd seama de proprietatea 3° a funcției de repartiție, rezultă:

$$F(x) = \int_{-\infty}^{x} f(x)\, dx. \tag{3.15}$$

Din relația (3.15) rezultă următoarea

Teoremă: Probabilitatea ca variabila aleatoare continuă X să ia valoare în intervalul (α, β) este egală cu integrala densității de repartiție pe intervalul (α, β):

$$P(\alpha \leqslant X < \beta) = \int_{\alpha}^{\beta} f(x)\, dx. \tag{3.16}$$

Această teoremă poate fi interpretată geometric.

Scanned with OKEN Scanner

Probabilitatea ca variabila aleatoare continuă X să ia o valoare în intervalul (α, β) este egală cu aria domeniului mărginit de axa x, de curba densității de repartiții $f(x)$ și de dreptele $x = \alpha$ și $x = \beta$ (vezi fig. 3.11).

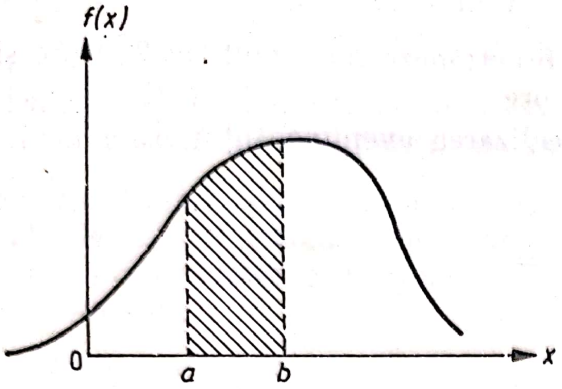


Fig. 3.11

Densitatea de repartiție are următoarele două proprietăți importante:
1°. Densitatea de repartiție este nenegativă:

$$f(x) \geqslant 0.$$

2°. Există relația:

$$\int_{-\infty}^{+\infty} f(x)\ dx = 1.$$

Aplicația 4. Viața medie a unui tip de lămpi de radio, în ore, are următoarea densitate de repartiție:

$$f(x) = \begin{cases} \frac{100}{x^2}, & \text{dacă } x \geqslant 100 \\ 0, & \text{dacă } x < 100. \end{cases}$$

Să se calculeze:
a) probabilitatea ca o lampă să nu fie înlocuită în primele 150 ore de funcționare;
b) probabilitatea ca din trei lămpi ale unui aparat de radio să nu se înlocuiască nici una în primele 150 ore de funcționare.
c) știind că o lampă funcționează mai mult de 200 ore, să se determine probabilitatea ca ea să nu funcționeze mai mult de 300 ore.

Soluție:
a) Probabilitatea ca lampa să funcționeze mai mult de 150 ore:

$$P(X > 150) = \int_{150}^{\infty} \frac{100}{x^2}\ dx = \left[-\frac{100}{x}\right]_{150}^{\infty} = \frac{100}{150} = \frac{2}{3}.$$

b) Fie A, B și C, evenimentele ca cele 3 lămpi să funcționeze mai mult de 150 ore. Probabilitățile lor sînt:

$$P(A) = P(B) = P(C) = \frac{2}{3}.$$

Scanned with OKEN Scanner

Evenimentele A, B, C fiind independente, probabilitatea producerii lor simultane este egală cu probabilitatea intersecției lor:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}.$$

c) Fie A evenimentul ca o lampă să funcționeze mai mult de 200 ore și B ca ea să funcționeze mai puțin de 300 ore.

Evenimentul B este condiționat de realizarea evenimentului A.

Avem deci:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(200 < X < 300)}{P(X > 200)} =$$

$$= \frac{\left[-\frac{100}{x}\right]_{200}^{300}}{\left[-\frac{100}{x}\right]_{200}^{\infty}} = \frac{\frac{1}{2} - \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

3.3 VALORI TIPICE ALE VARIABILELOR ALEATOARE

În paragraful precedent s-a arătat că o variabilă aleatoare poate fi caracterizată prin repartiția ei. Uneori însă pentru a caracteriza diferite variabile aleatoare trebuie să recurgem la mărimi de un fel deosebit, numite *valori tipice*, asociate acestora. Printre acestea se numără, în primul rînd, valoarea medie, dispersia, abaterea medie pătratică, valori medii de ordin superior (momente), mediana, modul etc.

3.3.1 Valori tipice ale variabilelor aleatoare discrete

1°. *Valoarea medie.* Printre valorile tipice ale unei variabile aleatoare un loc important îl ocupă valoarea medie [6].

Prin definiție, *valoarea medie a unei variabile aleatoare* cu repartiția:

$$X : \begin{pmatrix} x_1 & x_2 \dots x_n \\ p_1 & p_2 \dots p_n \end{pmatrix}$$

este egală cu suma produselor dintre valorile pe care le poate lua X și probabilitățile respective.

Vom nota valoarea medie a variabilei cu $M(X)$ sau m și uneori $\overline{X}$ (cînd este vorba de valori observate empiric); adoptînd prima notație putem scrie:

$$M(X) = p_1x_1 + p_2x_2 + \dots + p_nx_n = \sum_{i=1}^{n} p_ix_i. \qquad (3.17)$$

[6] Un termen echivalent, mai vechi, este speranța matematică.

Scanned with OKEN Scanner

Aplicația 5. Dacă repartiția X este:

$$X:\begin{pmatrix} 3 & 5 & 2 \\ 0{,}4 & 0{,}1 & 0{,}5 \end{pmatrix},$$

atunci:

$$M(X) = 3 \cdot 0{,}4 + 5 \cdot 0{,}1 + 2 \cdot 0{,}5 = 2{,}7.$$

Aplicația 6. Probabilitatea unui eveniment A este egală cu p. Să se calculeze media numărului de apariții ale evenimentului A într-un experiment.

Soluție. Variabila X — numărul de apariții ale evenimentului A — poate lua valorile $x_1 = 1$ (dacă evenimentul A apare) cu probabilitatea p și $x_2 = 0$ (dacă evenimentul A nu apare), cu probabilitatea $q = 1 - p$. Prin urmare, în acest caz:

$$M(X) = 1 \cdot p + 0 \cdot q = p. \tag{3.18}$$

Deci, media numărului de apariții ale unui eveniment, într-un singur experiment, este egală cu probabilitatea acestui eveniment.
Vom folosi mai departe acest rezultat.

Între valoarea medie a unei variabile aleatoare X și media aritmetică a valorilor observate ale lui X în n experimente independente există o anumită legătură.

Să presupunem că în cele n experimente independente, variabila X a luat de n_1 ori valoarea x_1, de n_2 ori valoarea x_2 ..., de n_m ori valoarea x_m, astfel încît $n_1 + n_2 + \ldots + n_m = n$. Este evident că suma tuturor valorilor luate de X este:

$$x_1 n_1 + x_2 n_2 + \ldots + x_m n_m.$$

Pentru a obține media aritmetică ($\overline{X}$) a valorilor luate de X trebuie să împărțim suma obținută la numărul total de experimente. Astfel avem:

$$\overline{X} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_m n_m}{n}$$

sau:

$$\overline{X} = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \ldots + x_m \frac{n_n}{n}.$$

După cum se vede, $\frac{n_1}{n}$ este frecvența relativă a valorii x_1, $\frac{n_2}{n}$ — frecvența relativă a valorii x_2 etc. Dacă numărul n al experimentelor este mare, frecvența relativă se apropie de probabilitate:

$$\frac{n_1}{n} \approx p_1, \frac{n_2}{n} \approx p_2, \ldots, \frac{n_n}{n} \approx p_m.$$

Înlocuind frecvențele relative prin probabilitățile corespunzătoare, obținem egalitatea aproximativă:

$$\overline{X} \approx x_1 p_1 + x_2 p_2 + \ldots + x_m p_m$$

adică:

$$\overline{X} \approx M(X). \tag{3.19}$$

Scanned with OKEN Scanner

Prin urmare, media variabilei aleatoare X este aproximativ egală cu media aritmetică a valorilor observate ale lui X, dacă numărul cazurilor observate este foarte mare.

Proprietăţile mediei. Vom prezenta — fără demonstraţii — proprietăţile mediei:

Prima proprietate. Media unei constante a este egală cu a:

$$M(a) = a.$$

Într-adevăr, constanta a poate fi considerată ca o variabilă aleatoare discretă care ia o singură valoare a cu probabilitatea $p = 1$. Prin urmare:

$$M(a) = a \cdot 1 = a. \tag{3.20}$$

A doua proprietate. Media produsului dintre o constantă a şi o variabilă aleatoare X este egală cu produsul dintre a şi media lui X:

$$M(a \cdot X) = a \cdot M(X).$$

Şi această proprietate este evidentă.

A treia proprietate. Media sumei a două variabile este egală cu suma mediilor celor două variabile aleatoare:

$$M(X + Y) = M(X) + M(Y). \tag{3.21}$$

Această proprietate este valabilă pentru un număr oarecare de variabile aleatoare:

$$M(X + Y + \ldots + Z) = M(X) + M(Y) + \ldots + M(Z). \tag{3.22}$$

Deci *media unei sume finite de variabile aleatoare este egală cu suma mediilor acestor variabile.*

Aplicaţia 7. Se aruncă două zaruri. Să se găsească media sumei numărului de puncte obţinute.

Soluţie: Să notăm cu X numărul de puncte ce se obţin prin aruncarea primului zar şi cu Y numărul de puncte ce se obţin prin aruncarea celui de-al doilea zar. Este evident că X şi Y sînt variabile aleatoare ce pot lua valorile 1, 2, 3, 4, 5 şi 6. Dacă zarurile sînt „ideale“, atunci X şi Y au fiecare repartiţia:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Prin urmare:

$$M(X) = M(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

şi

$$M(X + Y) = M(X) + M(Y) = \frac{7}{2} + \frac{7}{2} = 7.$$

A patra proprietate: Media produsului a două variabile aleatoare independente este egală cu produsul mediilor celor două variabile aleatoare:

$$M(X \cdot Y) = M(X) \cdot M(Y). \tag{3.23}$$

Scanned with OKEN Scanner

Această proprietate se menține și pentru un număr oarecare de variabile aleatoare independente: *media unui produs finit de variabile aleatoare independente este egală cu produsul mediilor variabilelor aleatoare*

$$M(X \cdot Y \cdots Z) = M(X) \cdot M(Y) \cdots M(Z). \tag{3.24}$$

Proprietățile a treia și a patra au o mare importanță deoarece prin aplicarea lor se obține mai simplu valoarea medie în diferite cazuri în care calculul direct este mai dificil.

Aplicația 8. Ne punem problema de a calcula valoarea medie a numărului de bile albe ieșite în n extracții succesive dintr-o urnă Bernoulli de compoziție dată, punînd de fiecare dată bila extrasă înapoi în urnă.

Soluție. Să notăm cu X_i variabila care dă numărul de bile albe ieșite în extracția de rang i. Deoarece în această extracție poate ieși o bilă albă sau niciuna, avem repartiția

$$X_i : \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}.$$

Numărul total X de bile ieșite în cele n extracții se obține însumînd variabilele X, adică:

$$X = X_1 + X_2 + \ldots + X_n,$$

de unde, aplicînd proprietatea a treia a sumei de variabile aleatoare, obținem:

$$M(X) = \sum_{i=1}^{n} M(X_i).$$

Însă

$$M(X_i) = p \cdot 1 + q \cdot 0 = p;$$

deci

$$M(X) = n \cdot p. \tag{3.25}$$

Acest rezultat îl vom întîlni la prezentarea valorilor tipice ale repartiției binomiale.

Aplicația 9. Se aruncă două zaruri. Care este valoarea medie a produsului numerelor de puncte obținute pe fața fiecărui zar?

Soluție. Fie X și Y variabilele aleatoare care ne dau, respectiv, numărul de puncte obținut pentru fiecare zar.
Avem:

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

$$Y : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

X și Y sînt variabile independente.

Scanned with OKEN Scanner

Prin urmare:

$$M(XY) = M(X) \cdot M(Y).$$

A cincea proprietate. În cursul procesului său de variaţie, o variabilă aleatoare X oscilează în jurul valorii sale medii.

Să notăm pentru uşurinţa scrierii $M(X) = m$ şi fie x_α cea mai mică dintre valorile $x_1, x_2 \ldots x_n$ şi x_β cea mai mare dintre aceste valori. Este evident că înlocuind în (3.17) valorile $x_1, x_2 \ldots x_n$ prin valoarea cea mai mică x_α, obţinem:

$$m \geqslant (p_1 + p_2 + \ldots + p_n)\, x_\alpha = x_\alpha.$$

În mod analog $m \leqslant x_\beta$. Avem deci dubla inegalitate

$$x_\alpha \leqslant m \leqslant x_\beta. \tag{3.26}$$

A şasea proprietate. Valoarea medie a abaterii faţă de media sa este egală cu zero.

Fie X o variabilă şi $M(X)$ media sa. Să considerăm variabila aleatoare $X - M(X)$ care se numeşte *abaterea variabilei aleatoare faţă de media sa (pe scurt — abatere).*

Dacă X are repartiţia:

$$X : \begin{pmatrix} x_1, & x_2, \ldots x_n \\ p_1, & p_2, \ldots p_n \end{pmatrix},$$

atunci abaterea $X - M(X)$ are repartiţia:

$$X - M(X) : \begin{pmatrix} x_1 - M(X) & x_2 - M(X) \ldots x_n - M(X) \\ p_1 & p_2 \quad \ldots \quad p_n \end{pmatrix}.$$

Se vede că:

$$\begin{aligned} M[X - M(x)] &= [x_1 - M(X)]\, p_1 + \ldots + [x_n - M(X)]\, p_n = \\ &= (x_1 p_1 + x_2 p_2 + \ldots + x_n p_n) - M(X)\,[p_1 + p_2 + \ldots + p_n] = \\ &= M(X) - M(X) = 0. \end{aligned} \tag{3.27}$$

Această proprietate ne pune în evidenţă faptul că pentru a măsura împrăştierea unei variabile aleatoare va trebui să recurgem la un alt indicator decît cel al mediei abaterilor de la medie.

Aplicaţia 10. Valoarea medie a numărului de puncte ieşite la aruncarea unui zar este egală cu:

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5.$$

Pentru verificarea proprietăţii a şasea, vom calcula $M\,[X - M(x)]$. Din repartiţia:

$$X - M(X) : \begin{pmatrix} -2{,}5 & 1{,}5 & 0{,}5 & 0{,}5 & 1{,}5 & 2{,}5 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

rezultă că $M[X - M(X)] = 0$.

Scanned with OKEN Scanner

2°. *Dispersia.* În teoria variabilelor aleatoare, această valoare tipică este fundamentală. Pentru a sesiza sensul ei cît mai clar cu putință vom porni de la următoarea aplicație:

Aplicația 11. Variabilele X și Y au următoarele repartiții:

$$X:\begin{pmatrix} -0{,}1 & 0{,}1 \\ 0{,}5 & 0{,}5 \end{pmatrix}, \quad Y:\begin{pmatrix} -1\,000 & 1\,000 \\ 0{,}5 & 0{,}5 \end{pmatrix}.$$

Să se calculeze mediile acestor variabile aleatoare și să se interpreteze rezultatul obținut:

$$M(X) = -0{,}1 \cdot 0{,}5 + 0{,}1 \cdot 0{,}5 = 0$$

$$M(Y) = -1\,000 \cdot 0{,}5 + 1\,000 \cdot 0{,}5 = 0.$$

Din acest exemplu se vede că valorile variabilei aleatoare X nu diferă mult de medie, pe cînd valorile lui Y diferă foarte mult. Aceasta înseamnă că pentru a caracteriza *împrăștierea* valorilor unei variabile aleatoare nu putem folosi valoarea medie, ci trebuie să recurgem la un alt indicator, cunoscut sub denumirea de *dispersie.*

Prin definiție, dispersia este valoarea medie a pătratului abaterii $X - M(X)$. O vom nota cu $D^2(X)$, definind-o deci astfel:

$$D^2(X) = M[X - M(X)]^2. \tag{3.28}$$

Aplicația 12. Ne propunem să calculăm dispersia variabilei aleatoare din aplicația 5. Am văzut că media variabilei era 2,7. Pentru a calcula dispersia vom considera repartiția abaterii ridicată la pătrat:

$$M[X - M(X)]^2 : \begin{pmatrix} (3-2{,}7)^2 & (5-2{,}7)^2 & (2-2{,}7)^2 \\ 0{,}4 & 0{,}1 & 0{,}5 \end{pmatrix};$$

valoarea medie a acestei variabile — dispersia — o calculăm imediat:

$$M[X - M(X)]^2 = 0{,}3^2 \cdot 0{,}4 + 2{,}3^2 \cdot 0{,}1 + (-0{,}7)^2 \cdot 0{,}5 =$$

$$= 0{,}036 + 0{,}529 + 0{,}245 = 0{,}810.$$

Dispersia se bucură de mai multe proprietăți importante.

Prima proprietate. Dispersia unei constante a este egală cu zero:

$$D^2(a) = 0.$$

Această proprietate este intuitiv evidentă, întrucît o mărime constantă nu are nici o împrăștiere.

De altfel, prin definiție, avem:

$$D^2(a) = M[a - M(a)]^2.$$

Folosind prima proprietate a mediei, obținem:

$$D^2(a) = M[(a-a)^2] = M(0) = 0. \tag{3.29}$$

Scanned with OKEN Scanner

A doua proprietate. Dispersia produsului dintre o constantă și o variabilă aleatoare X este egală cu produsul dintre pătratul constantei și dispresia variabilei X:

$$D^2(aX) = a^2D^2(X).$$

A treia proprietate. Dispersia sumei a două variabile aleatoare independente este egală cu suma dispersiilor celor două variabile aleatoare:

$$D^2(X + Y) = D^2(X) + D^2(Y).$$

Din această proprietate rezultă și două consecințe:

Consecința 1. Dispersia unei sume finite de variabile aleatoare independente este egală cu suma dispersiilor variabilelor aleatoare:

$$D^2(X + Y + \cdots + Z) = D^2(X) + D^2(Y) + \ldots D^2(Z). \tag{3.30}$$

Consecința 2. Dispersia sumei dintre o constantă a și o variabilă aleatoare X este egală cu dispersia lui X:

$$D^2(a + X) = D^2(X).$$

Este evident că a și X sînt independente. Folosind proprietățile cunoscute ale dispersiei, avem:

$$D^2(a + X) = D^2(a) + D^2(X) = D^2(X). \tag{3.31}$$

Aplicația 13. Dintr-o urnă de tip Bernoulli cu două stări — proporția bilelor albe și negre este cunoscută — se fac n extracții punîndu-se bila extrasă înapoi în urnă, după fiecare extracție. Să se calculeze care este dispersia numărului de bile albe extrase din urnă.

Soluție. După cum știm putem scrie egalitatea:

$$X = X_1 + X_2 + \ldots + X_n,$$

unde X este numărul total de bile albe extrase în n extracții succesive, iar variabilele X_1, X_2, ... X_n sînt independente și au aceeași repartiție:

$$X_i : \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}.$$

În virtutea consecinței 1 a proprietății a treia a dispersiei putem scrie:

$$D^2(X) = D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n).$$

Cum fiecare din variabilele X_1^2, X_2^2, ..., X_n^2 are repartiția

$$\begin{pmatrix} 1^2 & 0^2 \\ p & q \end{pmatrix}$$

avem:

$$D^2(X) = M(X_i^2) - M(X_i)^2 = 1^2p + 0^2q - (1 \cdot p + 0 \cdot q)^2 =$$
$$= p - p^2 = p(1 - p) = pq; \qquad 1 \leqslant i \leqslant n$$

Scanned with OKEN Scanner

Prin urmare:

$$D^2(X) = \underbrace{pq + pq + \ldots + pq}_{\text{de } n \text{ ori}} = npq. \tag{3.32}$$

Acest rezultat este foarte important; el definește dispersia unei repartiții binominale de care ne vom mai ocupa.

3°. *Abaterea medie pătratică.* Se numește abatere medie pătratică a unei variabile aleatoare rădăcina pătrată din dispersia variabilei aleatoare. Prin urmare, abaterea medie pătratică a unei variabile aleatoare X este:

$$D(X) = \sqrt{D^2(X)} = \sqrt{\sigma^2} = \sigma. \tag{3.33}$$

Folosind proprietățile cunoscute ale dispersiei putem afirma că abaterea medie pătratică a unei sume de variabile aleatoare independente este egală cu rădăcina pătrată din suma pătratelor abaterilor medii pătratice ale variabilelor aleatoare. În cazul aplicației 13, valoarea abaterii medii pătratice este:

$$\sigma = \sqrt{npq}.$$

Spre deosebire de dispersie, abaterea medie pătratică se exprimă în aceleași unități de măsură ca și variabila aleatoare.

4°. *Valori tipice de ordin superior (momente).* Adeseori în rezolvarea problemelor practice se dovedește a fi util calculul valorilor tipice de ordin superior ale variabilelor aleatoare. Este vorba de așa-numitele *momente* ale variabilei aleatoare.

Momentul de ordinul k al unei variabile aleatoare X este media variabilei aleatoare X^k. Dacă notăm cu ν_k un asemenea moment putem scrie:

$$\nu_k = M(X^k). \tag{3.34}$$

În particular:

$$\nu_1 = M(X)$$

$$\nu_2 = M(X^2)$$

$$\nu_3 = M(X^3).$$

Momentul centrat de ordinul k al variabilei aleatoare X este definit de media variabilei aleatoare $[X - M(X)]^k$. Dacă notăm cu μ_k aceste momente, putem scrie:

$$\mu_k = M[X - M(X)]^k. \tag{3.35}$$

În particular:

$$\mu_1 = M[X - M(X)] = 0$$

$$\mu_2 = M[X - M(X)]^2 = D^2(X) = \sigma^2. \tag{3.36}$$

Relațiile dintre momentele obișnuite și momentele centrate sînt ușor de dedus.

Scanned with OKEN Scanner

Prin definiție:

$$\mu_k = M(X - \nu_1)^k.$$

Însă:

$$(X - \nu_1)^k = X^k - C_k^1 \nu_1^1 X^{k-1} + C_k^2 \nu_1^2 X^{k-2} + \ldots (-1)^k \nu_1^k.$$

Prin urmare:

$$\mu_k = M[X^k - C_k^1 \nu_1 X^{k-1} + C_k^2 \nu_1^2 X^{k-2} + \ldots + (-1)^k \nu_1^k] =$$
$$= M(X^k) - C_k^1 \nu_1 M(X^{k-1}) + C_k^2 \nu_1^2 M(X^{k-2}) + \ldots + (-1)^k \nu_1^k =$$
$$= \nu_k - C_k^1 \nu_1 \nu_{k-1} + C_k^2 \nu_1^2 \nu_{k-2} + \ldots + (-1)^k \nu_1^k. \qquad (3.37)$$

Pe baza formulei (3.37), făcînd $k = 2$, obținem:

$$\mu_2 = \nu_2 - \nu_1^2,$$

adică dispersia este egală cu diferența dintre momentul de ordinul doi și momentul de primul ordin, ridicat la pătrat.

Făcînd succesiv, $k = 3, 4 \ldots$ rezultă:

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 - 2\nu_1^3$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4$$

.....

Putem considera și abaterea de la o valoare a arbitrar aleasă a variabilei aleatoare X; *obținem momente* de forma:

$$\lambda_k = M(X - a)^k$$

a căror utilitate o vom vedea cu prilejul expunerilor cu privire la repartițiile empirice.

3.3.2 Valori tipice ale variabilelor aleatoare continue. Valori tipice de poziție

Repartițiile continue sînt definite prin densitatea de probabilitate respectivă $f(x)$, care îndeplinește condițiile arătate:

$$f(x) \geq 0; \quad \int_{-\infty}^{+\infty} f(x)\, dx = 1.$$

Am văzut că în cazul repartițiilor discrete valorile tipice se definesc prin sume; în cazul repartițiilor continue, sumele se înlocuiesc prin integrale.

Prin analogie cu definițiile date valorii medii și dispersiei pentru variabile aleatoare discrete, putem formula definiții pentru aceste valori tipice și în cazul variabilelor continue.

Scanned with OKEN Scanner

Media unei variabile aleatoare continue X care ia valori în intervalul (x_1, x_2) şi a cărei densitate de repartiţie este $f(x)$ este integrala definită de la x_1 la x_2 a produsului $xf(x)$:

$$M(X) = \int_{x_1}^{x_2} xf(x)\,dx. \tag{3.38}$$

Dispersia variabilei aleatoare continue X în intervalul (x_1, x_2) este media pătratului abaterii lui X:

$$D^2(X) = \int_{x_1}^{x_2} [X - M(X)]^2 f(x)\,dx. \tag{3.39}$$

Momentul de ordinul k este:

$$\nu_k = \int_{-\infty}^{+\infty} x^k f(x)\,dx, \tag{3.40}$$

iar momentul centrat μ_k este:

$$\mu_k = \int_{-\infty}^{+\infty} (x - \nu_1)^k f(x)\,dx. \tag{3.41}$$

Toate proprietăţile momentelor pe care le-am stabilit pentru repartiţiile discrete se păstrează şi pentru repartiţiile continue.

Aplicaţia 14. Se dă densitatea de repartiţie $f(x)$:

$$f(x) = \begin{cases} (1 - x^2), & \text{pentru } -1 \leqslant x \leqslant 1 \\ 0, & \text{pentru } x > 1 \text{ şi } x < -1. \end{cases}$$

Să se calculeze valoarea medie şi dispersia repartiţiei corespunzătoare. Verificăm că $f(x)$ este într-adevăr o densitate de repartiţie observînd că:

$$f(x) \geqslant 0 \qquad \text{pentru } x \in [-1, 1]$$

şi

$$\frac{3}{4}\int_{-1}^{1} (1 - x^2)\,dx = \frac{3}{4}\left[x - \frac{x^3}{3}\right]_{-1}^{1} = 1.$$

Valorile tipice sînt:

$$\nu_1 = \frac{3}{4}\int_{-1}^{1} x(1 - x^2)\,dx = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{1} = \frac{3}{8}$$

Scanned with OKEN Scanner

$$\nu_2 = \frac{3}{4}\int_{-1}^{1} x^2(1-x^2)\,dx = \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1} = \frac{1}{5}$$

$$D^2(x) = \nu_2 - \nu_1^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320}.$$

Deci valoarea medie este egală cu $\frac{3}{8}$, iar dispersia egală cu $\frac{19}{320}$.

Valori tipice de poziţie—moda (modul) şi mediana. În afară de valoarea medie care caracterizează *poziţia* unei variabile aleatoare pe o axă, indicînd orientativ o valoare centrală în jurul căreia se grupează toate valorile posibile ale variabilei, în practică se mai folosesc şi alte două valori de poziţie: moda (modul), pe care o vom nota cu *Mo* şi mediana, pe care o vom nota cu *Me*.

a) Prin *modă (Mo)* se înţelege o valoare a variabilei aleatoare care este cea mai probabilă.

Dacă poligonul de repartiţie are două sau mai multe maxime, repartiţia se numeşte *plurimodală*.

Există repartiţii care înregistrează şi valori minime ale densităţii de repartiţie între cele două extreme; asemenea repartiţii se numesc *antimodale*.

b) *Mediana* variabilei aleatoare (*Me*) este tot o valoare tipică care se calculează mai ales pentru variabile aleatoare continue, deşi, formal, ea poate fi calculată şi pentru variabile discrete.

Fie X o variabilă aleatoare a cărei funcţie de repartiţie este $F(x)$.

Vom spune că numărul *Me* este mediana variabilei X dacă:

$$F(x) \leqslant \frac{1}{2}, \text{ pentru } x \leqslant Me$$

$$F(x) \geqslant \frac{1}{2}, \text{ pentru } x > Me. \tag{3.42}$$

Aplicaţia 15. Se aruncă două zaruri şi se consideră variabila aleatoare X constînd din suma punctelor obţinute pe cele două feţe ale zarurilor. Să se calculeze mediana lui X.

Soluţie:

$$X: \begin{Bmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{Bmatrix}.$$

Observăm că:

$$F(7) = P(X < 7) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{15}{36} < \frac{1}{2}$$

$$F(8) = \frac{1}{36}\ \frac{2}{36}\ \frac{3}{36}\ \frac{4}{36}\ \frac{5}{36}\ \frac{6}{36} = \frac{21}{36} > \frac{1}{2}.$$

Scanned with OKEN Scanner

Funcția de repartiție, în cazul repartiției discrete — ca în această aplicație — este o funcție în trepte. În fig. 3.12. este dat graficul acestei funcții, în punctele 6, 7 și 8.

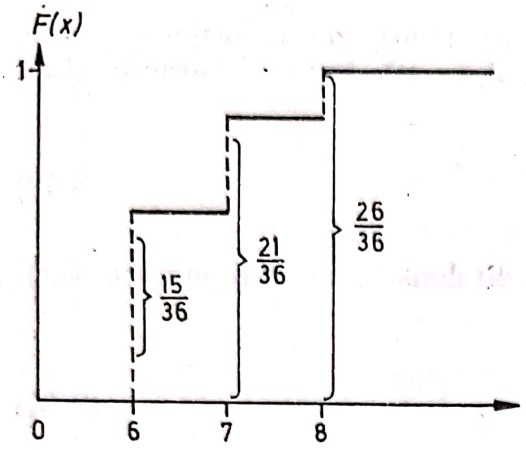


Fig. 3.12

În punctul $x = 7$, funcția de repartiție sare de la valoarea $\frac{15}{36}$ la $\frac{21}{36}$. Prin urmare:

$$F(x) < \frac{1}{2}, \text{ pentru } x \leqslant 7$$

$$F(x) > \frac{1}{2}, \text{ pentru } x > 7.$$

Rezultă $Me = 7$.

Aplicația 16. Se dă variabila aleatoare continuă X definită prin densitatea de repartiție:

$$f(x) = \begin{cases} x, & \text{pentru } x \in [0, \sqrt{2}] \\ 0, & \text{pentru } x \notin [0, \sqrt{2}]. \end{cases}$$

Să se calculeze mediana lui X.

Soluție: Funcția de repartiție este:

$$F(x) = \int_0^x t \, dt = \frac{x^2}{2}.$$

Pentru $x = 1$ rezultă:

$$F(1) = \frac{1}{2}.$$

Deci $Me = 1$.

c) *Funcție caracteristică.* Un concept de importanță fundamentală pentru cercetările moderne de teorie a probabilităților este conceptul de *funcție caracteristică.* Două legi de repartiție care au aceeași funcție caracteristică sînt *identice*, de unde provine și numele acestei funcții.

Scanned with OKEN Scanner

Unei variabile aleatoare X, i se poate asocia o funcție cu valori complexe, definită ca *valoare medie a expresiei* e^{ixt}, unde t este o variabilă reală, iar i o unitate imaginară. O asemenea funcție se numește funcție caracteristică și o vom nota cu $\varphi(t)$.

Să considerăm o variabilă aleatoare X care poate lua valorile $x_j (j = 1, 2 ..., n)$, cu probabilitățile respective p_j $(j=1, 2, ..., n)$. Funcția caracteristică corespunzătoare este următoarea:

$$\varphi(t) = \sum_{j=1}^{n} p_j e^{itx_j}, \quad \sum_{j=1}^{n} p_j = 1, \quad p_j \geqslant 0. \tag{3.43}$$

În cazul unei variabile aleatoare continue, cu densitatea de repartiție $f(x)$, definim funcția caracteristică astfel:

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx. \tag{3.44}$$

Funcția caracteristică este un instrument matematic foarte puternic pentru studiul variabilelor aleatoare. De pildă, prin derivări succesive ale funcției caracteristice putem determina valorile momentelor, bineînțeles în cazul cînd acestea sînt finite.[7]

3.4 REPARTIȚII ALE UNOR VARIABILE ALEATOARE

3.4.1 Repartiții ale unor variabile aleatoare de tip discret

Repartiția binomială. La prezentarea acestui tip de repartiție a probabilităților vom folosi unele elemente referitoare la modelul teoretic al urnei lui Bernoulli cu două stări.

Să presupunem deci că se fac n experimente independente după modelul urnei Bernoulli.

În fiecare experiment, probabilitatea de apariție a unui eveniment A este constantă și egală cu p, iar probabilitatea evenimentului non A este $q = 1 - p$. Este evident că numărul de realizări ale evenimentului A în cele n experimente constituie o variabilă aleatoare X. Valorile pe care le poate lua această variabilă aleatoare sînt 0, 1, 2 ..., n. Într-adevăr în cele n experimente, evenimentul A poate să nu se realizeze niciodată sau poate să se realizeze o dată, de două ori, ..., de n ori.

Probabilitățile valorilor X se găsesc aplicînd modelul lui Bernoulli care ne dă probabilitatea de realizare a evenimentului A de α ori în n experimente; modelul reprezintă termenul general al dezvoltării binomului lui Newton de forma:

$$(p + q)^n = C_n^n p^n q^0 + C_n^{n-1} p^{n-1} q^{n-(n-1)} + ... + C_n^{\alpha} p^{\alpha} q^{n-\alpha} + ... + \\ + C_n^0 p^0 q^{n-0}. \tag{3.45}$$

[7] Vezi o aplicație a funcției caracteristice la calculul unor momente ale repartiției binomiale din paragraful următor (aplicația 17).

Scanned with OKEN Scanner

Primul termen al acestei dezvoltări ne dă probabilitatea ca variabila aleatoare X să ia valoarea n; al doilea termen, $C_n^{n-1}p^{n-1}q^{n-(n-1)}$, ne dă probabilitatea ca X să ia valoarea $n-1, \ldots$, ultimul termen q^n ne dă probabilitatea ca X să ia valoarea 0.

Aşadar, legea de repartiţie a lui X este:

$$X:\begin{pmatrix} n & n-1 & \ldots & \alpha & \ldots & 0 \\ p^n & C_n^n p^{n-1}q^{n-(n-1)} & \ldots & C_n^\alpha p^\alpha q^{n-\alpha} & \ldots & C_n^0 p^0 q^{n-0} \end{pmatrix}. \tag{3.46}$$

Această lege de repartiţie a probabilităţilor se numeşte *binomială*.

Pentru repartiţia binomială, valoarea medie şi dispersia sînt date de relaţiile:

$$m = M(X) = \sum_{\alpha=0}^{n} \alpha C_n^\alpha p^\alpha q^{n-\alpha} = np \tag{3.47}$$

$$\sigma^2 = D^2(X) = npq. \tag{3.48}$$

Formulele (3.47) şi 3.48) le-am ilustrat prin aplicaţii concrete în cele precedente. Vom prezenta acum posibilitatea de a defini şi alte momente ale repartiţiei binomiale, cu ajutorul funcţiei caracteristice a acestei repartiţii.

Dacă o variabilă aleatoare X se supune unei repartiţii binomiale, funcţia ei caracteristică, prin definiţie, este următoarea:

$$\varphi(t) = M(e^{it\alpha}) = \sum_{\alpha=0}^{n} e^{\alpha ti}(C_n^\alpha p^\alpha q^{n-\alpha}) = \sum_{\alpha=0}^{n} C_n^\alpha (pe^{ti})^\alpha q^{n-\alpha} = (pe^{ti} + q)^n. \tag{3.49}$$

Folosind teoremele cu privire la valorile medii ale sumei şi produsului de variabile independente — teoreme ce ne sînt de acum cunoscute — putem deduce formulele de calcul al diferitelor momente pentru repartiţiile binomiale.

Primele patru momente sînt ilustrate prin următoarele egalităţi:

$$M(X) = np;\ M(X^2) = n^2p^2 + npq;\ M(X^3) = n^3p^3 +$$
$$+ 3n^2p^2q + npq(1-2p);$$
$$M(X^4) = n^4p^4 + 6n^3p^3q + n^2p^2(7 - 18p + 11p^3) +$$
$$+ np(1 - 7p + 12p^2 - 6p^3).$$

Momentele centrate sînt:

$$M(X-np) = 0;\ M(X-np)^2 = npq;\ M(X-np)^3 = npq(q-p);$$
$$M(X-np)^4 = 3n^2p^2q^2 + npq(1-6pq). \tag{3.50}$$

Aceste momente le mai putem nota, respectiv: μ_0, μ_1, μ_2 ... Momentele obişnuite se calculează cu ajutorul funcţiei caracteristice.

Din formula (3.49) — calculînd prima derivată — obţinem:

$$\varphi'(t) = npie^{ti}(pe^{ti} + q)^{n-1};\quad \varphi'(0) = npi.$$

Scanned with OKEN Scanner

Se poate arăta că:

$$m_k = \frac{\varphi^{(k)}(0)}{i^k};$$

rezultă că:

$$m_1 = \frac{npi}{i} = np.$$

Prin derivări succesive se obţin şi celelalte momente.

Aplicaţia 17. Un sector municipal este servit de 10 centre de desfacere a laptelui, fiecare centru funcţionînd independent unul de celălalt. Probabilitatea ca oricare din aceste magazine să fie în imposibilitate de a satisface cererea populaţiei la lapte este de 0,1.

Numărul magazinelor aflate în situaţia de a nu putea acoperi cererea populaţiei constituie o variabilă aleatoare X.

Se cere rezolvarea următoarelor probleme:

1. Să se arate legea de repartiţie a acestei variabile aleatoare. 2. Să se calculeze valoarea funcţiei caracteristice. 3. Să se calculeze probabilitatea ca cel puţin 5 magazine să poată acoperi cererea de consum de lapte a populaţiei. 4. Să se calculeze primele patru momente centrate ale variabilei.

Soluţii:

1. Avem o repartiţie binomială cu $p = 0{,}1$; $q = 0{,}9$; $n = 10$. Posibilitatea ca α ($\alpha = 0, 1, 2, ..., 10$) magazine să fie în imposibilitate de a livra marfa este, după formula generală:

$$C_{10}^{\alpha}(0{,}1)^{\alpha}\,(0{,}9)^{10-\alpha}.$$

2. Aplicînd formula (3.49) rezultă:

$$\varphi(t) = [0{,}1\ e^{ti} + 0{,}9]^{10}.$$

3. Probabilitatea căutată este:

$$(0{,}1)^{10} + C_{10}^{1}(0{,}1)^{1}\,(0{,}9)^{9} + C_{10}^{2}(0{,}1)^{2}(0{,}9)^{8} + C_{10}^{3}(0{,}1)^{3}(0{,}9)^{7} +$$

$$+ C_{10}^{4}(0{,}1)^{4}\,(0{,}9)^{6}.$$

4. Aplicînd formulele din cazul general rezultă:

$$\mu_1 = 0;\ \mu_2 = npq = 0{,}9;\ \mu_3 = npq(q - p) = 0{,}72.$$

$$\mu_4 = 3n^2p^2q^2 + npq(1 - 6pq) = 1{,}24.$$

Repartiţie Poisson. Să considerăm n experimente independente, probabilitatea de realizare a unui eveniment A fiind constantă şi egală cu p.

Pentru determinarea probabilităţii ca în n experimente evenimentul A să se realizeze de α ori folosim formula cunoscută a lui Bernoulli. Dacă n este

Scanned with OKEN Scanner

mare şi valoarea probabilităţii p, foarte mică ($p \leqslant 0{,}1$) putem recurge la formula asimtotică a lui Poisson. Această formulă se obţine, admiţîndu-se ipoteza că produsul np păstrează o valoare constantă λ; $np = \lambda$. Aceasta înseamnă că valoarea medie a numărului de realizări ale evenimentului A, în diferite serii de experimente pentru valori diferite ale lui n, rămîne neschimbată.

Se poate demonstra, în acest caz, următoarea teoremă de limită:

$$P_n(\alpha;\lambda) = \lim_{n\to\infty} P_n(\alpha) = \frac{\lambda^\alpha}{\alpha!} e^{-\lambda}. \tag{3.51}$$

Această formulă defineşte *repartiţia Poisson*. Fiecărei valori întregi $\alpha = 0,1, \ldots, n-1, n \ldots$ îi corespunde o probabilitate $P(\alpha, \lambda)$.

Ea se scrie astfel:

$$X : \begin{pmatrix} 0 & 1 & 2 & \ldots & \alpha & \ldots \\ e^{-\lambda} & \lambda e^{-\lambda} & \frac{\lambda^2}{2!} e^{-\lambda} & \ldots & \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} & \ldots \end{pmatrix}.$$

Se vede că:

$$\sum_{\alpha=0}^{\infty} P(\alpha;\ \lambda) = \sum_{\alpha=0}^{\infty} \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} = e^{-\lambda} \sum_{\alpha=0}^{\infty} \frac{\lambda^\alpha}{\alpha!} = e^{-\lambda} \cdot e^{-\lambda} = 1.$$

Una din proprietăţile cele mai importante ale repartiţiei Poisson constă în faptul că media şi dispersia variabilei aleatoare X, care se supune acestei repartiţii, sînt egale, adică:

$$M(X) = \sum_{\alpha=0}^{\infty} \alpha \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} = \lambda; \quad D^2(X) = \lambda.$$

Această proprietate a repartiţiei Poisson se aplică adeseori în practică la verificarea ipotezelor statistice cu privire la repartiţia teoretică a datelor observate. De pildă, dacă media şi dispersia, calculate pe baza datelor observate, sînt egale sau foarte apropiate una de alta, avem suficiente temeiuri pentru a considera că repartiţia teoretică a variabilei studiate este de tip Poisson şi invers, dacă valorile numerice ale mediei şi dispersiei diferă sensibil, justeţea ipotezei cu privire la repartiţia Poisson trebuie pusă la îndoială.

Legea de repartiţie (3.51) poate fi scrisă şi în felul următor:

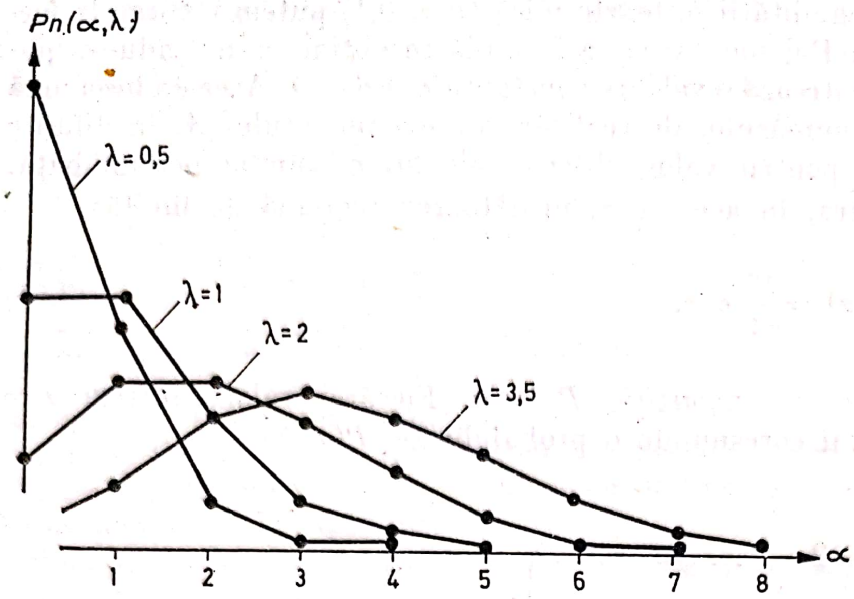
$$P(\alpha,\ np) = \frac{(np)^\alpha}{\alpha!} e^{-np}, \tag{3.52}$$

unde $\lambda = np$ reprezintă parametrul acelei repartiţii Poisson care înlocuieşte cu aproximaţie repartiţia binomială.

Repartiţia Poisson mai este denumită *legea evenimentelor rare*, termen folosit datorită proprietăţii sale de a aproxima o repartiţie binomială cînd numărul experimentelor (n) este foarte mare, iar probabilitatea de apariţie a

Scanned with OKEN Scanner

Fig. 3.13



evenimentului considerat (A) este foarte mică. În fig. 3.13 se dau forme ale curbelor de repartiție Poisson, considerîndu-se diferite valori ale lui λ.

În cazul unei variabile aleatoare care se supune legii de repartiție Poisson, funcția caracteristică este dată de expresia:

$$\varphi(t) = e^{-\lambda} \sum_{\alpha=0}^{\infty} \frac{\lambda^{\alpha}}{\alpha!} e^{-it\alpha} = e^{\lambda(e^{it}-1)}. \tag{3.53}$$

Aplicația 18. La o centrală telefonică se înregistrează, în medie, 240 chemări pe oră. Știind că numărul chemărilor telefonice în orice perioadă de timp se repartizează după legea Poisson, să se calculeze probabilitatea evenimentului ca în decurs de 2 minute să aibă loc exact 3 chemări.

Numărul mediu de chemări pe care centrala le primește în două minute este:

$$\lambda = \frac{2 \cdot 240}{60} = \frac{240}{30} = 8.$$

Legea de repartiție a variabilei aleatoare X, care reprezintă numărul de chemări potrivit (3.52), este următoarea:

$$P(3;8) = \frac{8^3}{3!} e^{-8}, \quad \alpha = 0,1, \ldots$$

dat fiind că în această aplicație avem:

$$\alpha = 3; \qquad \lambda = 8.$$

În tabele speciale găsim valoarea probabilității căutate [8].

$$P(3;8) = 0,0286.$$

[8] Vezi o serie de aplicații interesante ale legii de repartiție Poisson, în [83].

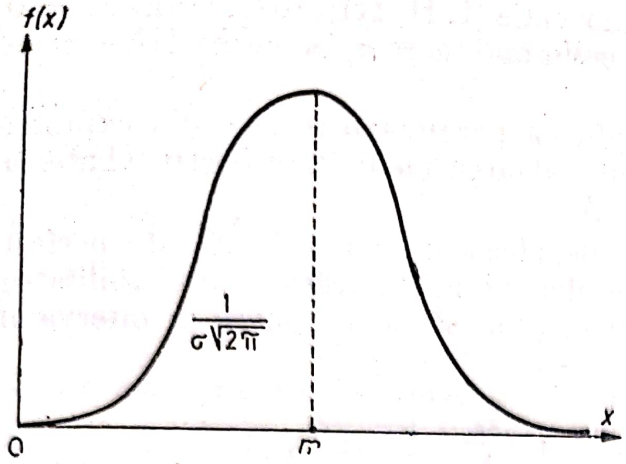


Fig. 3.14

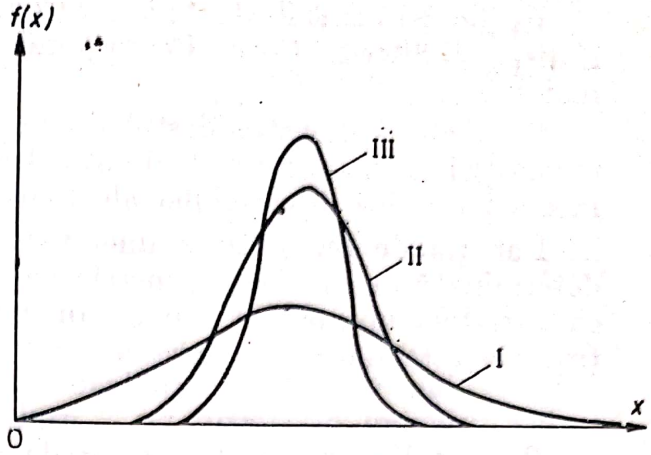


Fig. 3.15

Un eveniment rar ar fi, de pildă, un număr de 20 chemări în decurs de 2 minute, în condițiile centralei considerate.

$$P(20;\ 8) = \frac{8^{20}}{20!}\, e^{-8} = 0{,}0002.$$

3.4.2 Repartiții ale unor variabile aleatoare de tip continuu

Repartiția normală. Această repartiție, denumită adesea legea lui Gauss sau legea lui Laplace, este fundamentală în teoria probabilităților și în aplicațiile ei. Ea se întîlnește frecvent în studiul fenomenelor care au loc în natură și în viața social-economică. O caracteristică esențială a acestei legi — în raport cu celelalte repartiții — constă în aceea că este o lege limită către care tind numeroase legi de repartiție, în anumite condiții tipice.

Legea normală este definită prin următoarea densitate de repartiție:

$$f(x;\ m,\ \sigma) = \frac{1}{\sigma\cdot\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}. \tag{3.54}$$

Graficul funcției $f(x; m, \sigma)$ se numește curba normală (curba lui Gauss), cu parametrii m și σ (vezi fig. 3.14 și 3.15).

Curbele normale au cîteva particularități comune:

a) toate curbele normale admit cîte un punct maxim ($x = m$) și scad necontenit, simetric, la dreapta și la stînga acestuia, apropiindu-se necontenit de axa absciselor;

b) toate curbele normale sînt simetrice față de o paralelă cu axa ordonatelor dusă prin punctul de maxim $x = m$; ordonata maximă este egală cu $\frac{1}{\sigma\sqrt{2\pi}}$;

c) toate curbele normale au forma de clopot, avînd convexitatea îndreptată în sus, în apropierea punctului maxim.

În punctele $m - \sigma$ și $m + \sigma$, curbele normale își schimbă covexitatea și cu cît σ este mai mic cu atît clopotul este mai ascuțit, iar cu cît σ este mai mare cu atît clopotul este mai turtit.

Scanned with OKEN Scanner

În fig. 3.15 sînt ilustrate trei curbe normale (I, II, III), toate avînd $m = 0$ și dispersii diferite. Curba I corespunde celui mai mare σ, iar curba III — celui mai mic.

Se poate demonstra destul de simplu că parametrii m și σ din expresia repartiției normale nu sînt altceva decît valoarea medie și respectiv, abaterea medie pătratică a variabilei aleatoare X.

Funcția de repartiție a unei variabile aleatoare normale X este perfect determinată de media și dispersia variabilei. Se poate arăta că probabilitatea ca variabila normală X, cu parametrii m și σ, să ia o valoare în intervalul (x_1, x_2) este dată de expresia

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x; m; \sigma)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}\, dx. \tag{3.55}$$

Dacă facem schimbarea de variabilă $y = \frac{x-m}{\sigma}$, $x = \sigma y + m$ și ținem seama că pentru $X = x_1$, $y = \frac{x_1 - m}{\sigma}$, iar pentru $X = x_2$ avem $y = \frac{x_2 - m}{\sigma}$, obținem:

$$P(x_1 < X < x_2) = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{\frac{x_1-m}{\sigma}}^{\frac{x_2-m}{\sigma}} e^{-\frac{y}{2}}\, dy = \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\frac{x_2-m}{\sigma}} e^{-\frac{y}{2}}\, dy - \int_{-\infty}^{\frac{x_1-m}{\sigma}} e^{-\frac{y}{2}}\, dy \right).$$

Dacă considerăm funcția normală de repartiție cu parametrii 0 și 1 — numită funcția Laplace și notată cu $\Phi(x)$

$$\Phi(x) = \int_{-\infty}^{x} f(y; 0; 1)\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y}{2}} dy,$$

atunci putem scrie sub o altă formă probabilitatea ca variabila aleatoare X să fie cuprinsă într-un interval:

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - m}{\sigma}\right) - \Phi\left(\frac{x_1 - m}{\sigma}\right). \tag{3.56}$$

Această relație ne poate servi pentru determinarea probabilității ca abaterea variabilei aleatoare $X - m$ să satisfacă inegalitatea

$$|X - m| < \alpha.$$

Într-adevăr, această inegalitate este echivalentă cu dubla inegalitate

$$-\alpha < X - m < \alpha$$

sau

$$-\alpha + m < X < \alpha + m.$$

Scanned with OKEN Scanner

Deci, aplicînd relația (3.56), avem:

$$P(|X-m|<\alpha)=P(-\alpha+m<X<\alpha+m)=\Phi\left(\frac{\alpha}{\sigma}\right)-\Phi\left(-\frac{\alpha}{\sigma}\right).$$

Se vede ușor că

$$\Phi(-x)=1-\Phi(x);$$

și prin urmare:

$$P(|X-m|<\alpha)=2\Phi\left(\frac{\alpha}{\sigma}\right)-1. \tag{3.57}$$

Din (3.57) rezultă o regulă foarte importantă, cunoscută sub numele de *regula celor trei sigma*. Dacă o variabilă aleatoare X este normală, atunci, practic, abaterea sa — în valoare absolută — nu depășește întreitul valorii sale medii pătratice.

Să considerăm, de pildă, $\alpha=3\sigma$ și aplicînd formula (3.57) obținem:

$$P(|X-m|<3\sigma)=2\Phi(3)-1.$$

Dat fiind că valoarea funcției lui Laplace [9] $\Phi(3)=0{,}9987$, găsim imediat probabilitatea ca abaterea în valoare absolută a variabilei aleatoare X să depășească un interval egal cu 3σ. Această probabilitate se dovedește a fi foarte mică. Într-adevăr, dacă:

$$P(|X-m|<3\sigma)=2\cdot 0{,}9987-1=0{,}9974,$$

atunci

$$P(|X-m|>3\sigma)=1-0{,}9974=0{,}0026.$$

Aplicația 19. Abaterea de la diametrul standard a diametrului pieselor fabricate de un strung automat urmează o lege normală cu media 3 mm și abaterea medie pătratică 1. Să se determine probabilitatea ca diametrul piesei să aibă abateri, în valoare absolută, cuprinse în intervalul 2 și 5 mm

Probabilitatea căutată se determină aplicînd formula (3.56), fiind date $m=3$, $\sigma=1$, $x_1=2$ și $x_2=5$.

$$P(2<X<5)=\Phi\left(\frac{5-3}{1}\right)-\Phi\left(\frac{2-3}{1}\right)=\Phi(2)-\Phi(-1).$$

În tabele speciale găsim $\Phi(2)=0{,}9772$ și $\Phi(-1)=0{,}1587$. Deci

$$P(2<X<5)=0{,}9772-0{,}1587=0{,}8155.$$

Repartiția χ^2. Introdusă în știință de F. R. Helmert în 1876 și pusă în valoare de K. Pearson, repartiția χ^2 este o repartiție continuă, definită pe axa $(0,\infty)$. Densitatea ei de repartiție este:

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)}\,x^{\frac{s}{2}-1}e^{-\frac{x}{2}}, & \text{pentru } x\geqslant 0\\ 0, & \text{pentru } x<0.\end{cases} \tag{3.58}$$

[9] Valorile acestei funcții se găsesc în anexa 3.

Scanned with OKEN Scanner

În această formulă, $\Gamma(p)$ reprezintă funcţia lui Euler;

$$\Gamma(p) = \int_0^{\infty} x^{p-1} \cdot e^{-x} \cdot dx.$$

Se poate arăta, prin calcul, că $\Gamma(p) = (p-1)!$, pentru orice număr întreg p şi $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Această densitate de repartiţie defineşte funcţia de repartiţie χ^2 cu parametrii s şi σ. Parametrul s se numeşte *numărul gradelor de libertate*, înţelegîndu-se prin aceasta numărul variabilelor independente a căror variaţie nu suferă nici o restricţie.

De pildă, dacă este observat un sistem complet de s evenimente cu probabilităţile respective:

$$x_1, x_2, \ldots, x_s$$

$$p_1, p_2, \ldots, p_s, \quad \left(\sum_{i=1}^{s} p_i = 1\right), \ (i = 1,2, \ldots, s)$$

avem de fapt o repartiţie cu s valori componente; dacă se cunosc $s-1$ probabilităţi putem s-o găsim şi pe ultima. Sîntem în cazul unei variabile cu $s-1$ grade de libertate.

Vom preciza acum cîteva caracteristici ale repartiţiei χ^2. Fie $X_1, X_2, \ldots, X$ un şir de variabile aleatoare normale, normate, independente, adică $N(0,1)$. Densitatea de repartiţie a variabilelor X $(n = 1,2, \ldots)$ este

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Să considerăm expresia:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2 \tag{3.59}$$

Se demonstrează că variabila aleatoare χ^2 admite densitatea de repartiţie (5.58).

Valoarea medie şi dispersia variabilei χ^2 sînt date de formulele:

$$M(\chi^2) = n, \quad D^2(\chi^2) = 2n. \tag{3.60}$$

Curba de repartiţie a lui χ^2 (Fig. 3.16) este utilă în diferite aplicaţii practice foarte importante.

Probabilitatea P ca să obţinem pentru χ^2 o valoare mai mică decît un număr dat χ_0^2, adică,

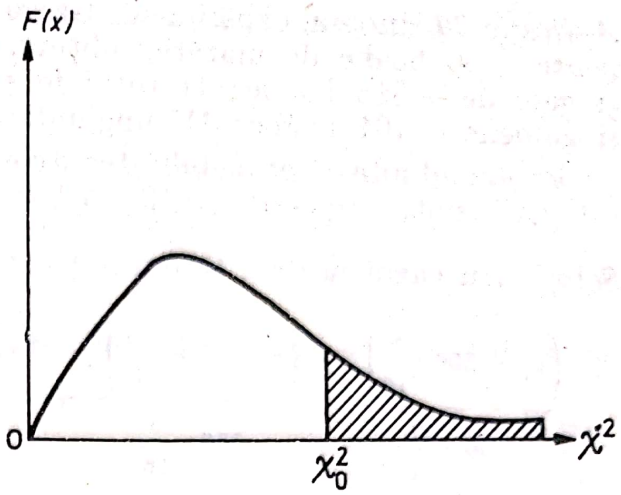
$$P = P(\chi^2 < \chi_0^2)$$

este egală cu aria nehaşurată în fig. 3.16. Expresia ei analitică este următoarea:

$$P = \frac{1}{2^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)} \int_0^{\chi_0^2} x^{\frac{s}{2}-1} dx = \int_0^{\chi_0^2} f(x)\, dx. \tag{3.61}$$

Scanned with OKEN Scanner

Fig. 3.16



Se întrebuinţează uneori în aplicaţii şi probabilitatea $P(\chi^2 > \chi_0^2)$. Valorile lui P sînt calculate în tabele [10]. Se constată din formula (3.61) că P variază în raport cu s — numărul de grade de libertate. Numeroasele aplicaţii ale lui χ^2 rezultă din proprietatea pe care o expunem mai jos. Considerăm un şir de variabile independente $Y_1, Y_2, \ldots Y_n$.

Fiecare variabilă $Y_i (i = 1, 2, 3, \ldots)$ poate lua valorile

$$a_1, a_2, \ldots, a_m,$$

cu probabilităţile respective

$$p_1, p_2, \ldots p_m, \quad \left(p_i \geqslant 0, \sum_{i=1}^{m} p_i = 1\right).$$

Considerînd grupul de variabile $Y_1, Y_2, \ldots Y_n$, să notăm prin $\alpha_i (i = 1,2, \ldots, m)$ numărul de cîte ori a fost luată valoarea a_i, de aceste m variabile. Avem:

$$\alpha_i \geqslant 0, \quad \sum_{i=1}^{m} \alpha_i = n.$$

Se observă că variabilele $Y_i (i = 1, 2, \ldots, n)$ se supun repartiţiei multinominale.

Probabilitatea ca să obţinem valorile α_i în n probe ştim că este:

$$\frac{n!}{\alpha_1!\ \alpha_2! \ldots \alpha_n!} p_1^{\alpha_1}\ p_2^{\alpha_2} \ldots p_m^{\alpha m}.$$

Considerăm, de asemenea, variabila aleatoare:

$$\frac{(\alpha_1 - np_1)^2}{np_1} + \frac{(\alpha_2 - np_2)^2}{np_2} + \ldots + \frac{(\alpha m - np_m)^2}{np_m}. \tag{3.62}$$

Variabila (3.62) urmează repartiţia χ^2 cu $(s - 1)$ grade de libertate. Această proprietate este importantă pentru că ne permite să verificăm unele ipoteze statistice.

[10] Vezi anexa 5.

Scanned with OKEN Scanner

Aplicaţia 20. Într-un experiment, Gregor Mendel a observat culoarea şi forma pentru 556 boabe de mazăre, obţinînd următoarele rezultate: I rotunde şi galbene — 315 boabe; II rotunde şi verzi — 108 boabe; III unghiulare şi galbene — 101 boabe; IV unghiulare şi verzi — 32 boabe.

Putem admite că probabilitatea de a obţine o boabă din una din cele patru categorii sînt, respectiv: 9/16; 3/16; 3/16; 1/16?

Soluţie .În cazul nostru, trebuie să considerăm variabila aleatoare:

$$\frac{\left(\alpha_1 - 556 \cdot \frac{9}{16}\right)^2}{556 \cdot \frac{9}{16}} + \frac{\left(\alpha_2 - 556 \cdot \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} + \frac{\left(\alpha_3 - 556 \cdot \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} + \frac{\left(\alpha_4 - 556 \cdot \frac{1}{16}\right)^2}{556 \cdot \frac{1}{16}}, \quad (3.63)$$

corespunzătoare variabilei aleatoare (3.62). Această variabilă aleatoare urmează legea χ^2 cu 3 grade de libertate.

Experimentul ne-a dat:

$$\alpha_1 = 315;\ \alpha_2 = 108;\ \alpha_3 = 101;\ \alpha_4 = 32,$$

ceea ce înseamnă că (3.63) a luat valoarea:

$$\frac{\left(315 - 556 \cdot \frac{9}{16}\right)^2}{556 \cdot \frac{9}{16}} + \frac{\left(108 - 556\ \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} + \frac{\left(101 - 556\ \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} +$$

$$\frac{\left(32 - 556 \cdot \frac{1}{16}\right)^2}{556 \cdot \frac{1}{16}} = 0{,}470$$

Din tabele găsim că pentru $s = 3$ avem probabilitatea:

$$P(\chi^2 > 0{,}352) = 0{,}95.$$

Rezultatul confirmă ipoteza făcută. Probabilitatea ca să obţinem o valoare mai mare decît 0,352 este foarte mare, fiind egală cu 0,95. Deci dintr-o singură probă este plauzibil să obţinem valoarea 0,470 care este mai mare decît 0,352.

Tot din tabele găsim, pentru $s = 3$, că

$$P(\chi^2 > 9{,}837) = 0{,}01.$$

Să presupunem că făcînd o altă experienţă ne-ar fi ieşit $\chi^2 = 10{,}48$. Cum $10{,}48 > 9{,}837$, înseamnă că dintr-un singur experiment ne-a ieşit un eveniment a cărui probabilitate de realizare este 0,01, ceea ce în mod normal nu se întîmplă. Nu avem deci un motiv pentru a afirma că ipoteza făcută este verosimilă. [11]

[11] Vezi şi aplicaţii ale repartiţiei lui χ^2 în [84] şi [85].

Scanned with OKEN Scanner

Repartiția Student [12]. Să considerăm un șir de variabile aleatoare normale, $N(0, \sigma)$ independente:

$$X_1, X_2, \ldots, X_n.$$

Densitatea de repartiție a fiecăreia din aceste variabile este prin urmare:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}.$$

Să mai considerăm acum o nouă variabilă aleatoare, notată de Student cu t:

$$t = \frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}, \tag{3.64}$$

în care: X de la numărător este o variabilă aleatoare independentă de șirul $X_1, X_2, \ldots, X_n$, normală, $N(0,\sigma)$, iar suma de la numitor este, evident, o variabilă aleatoare care urmează o lege de repartiție χ^2 cu parametrii s și σ.

Variabila t urmează o repartiție *continuă*, definită pe $(-\infty, +\infty)$, a cărei densitate de repartiție este dată de următoarea expresie:

$$f(t,s) = \frac{1}{\sqrt{s\pi}} \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{s}{2}\right)} \left(1 + \frac{t^2}{s}\right)^{-\frac{s+1}{2}}. \tag{3.65}$$

Numărul s reprezintă numărul gradelor de libertate ale repartiției t.

Funcția (3.65) fiind o densitate de repartiție, avem:

$$\frac{1}{\sqrt{s\pi}} \cdot \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{s}{2}\right)} \int_{-\infty}^{+\infty} \left(1 + \frac{t^2}{s}\right)^{-\frac{s+1}{2}} dt = 1. \tag{3.66}$$

Din cauza simetriei formulei (3.65) față de $t = 0$, toate momentele de ordin impar ale variabilei t sînt nule.

Se poate arăta deci că media și dispersia variabilei aleatoare a lui Student cu 1 grade de libertate sînt:

$$M(t) = 0$$

$$D^2(t) = \frac{s}{s-2}.$$

Momentele pare de un ordin mai mare decît s sînt infinite.

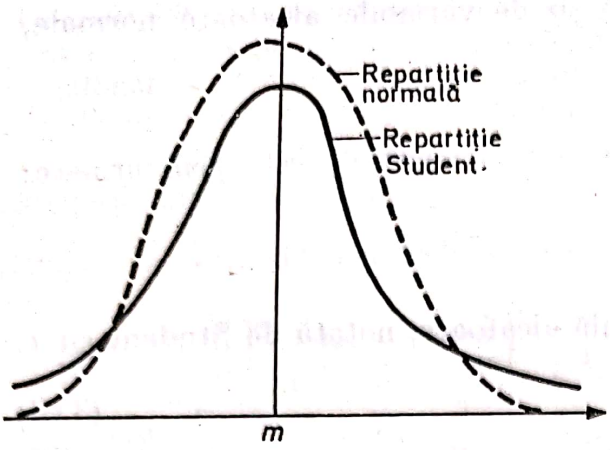
În fig. 3.17 este ilustrată curba teoretică a densității de repartiție a variabilei aleatoare t, în comparație cu curba normală. Se observă că ea este asemănătoare cu curba normală dar diferită de ea.

Cînd $s \to \infty$, funcția de repartiția a variabilei lui Student se apropie de funcția lui Laplace $\Phi(x)$.

[12] Student este pseudonimul ilustrului matematician W. Gosset cu care-și semna lucrările. Gosset a găsit primul această repartiție în 1908 și a arătat importanța ei pentru aplicațiile practice.

Scanned with OKEN Scanner

Fig. 3.17



Funcția de repartiție a variabilei aleatoare t este definită de expresia:

$$F(t, s) = \int_{-\infty}^{t} f(t, s)\, dt. \tag{3.67}$$

Studiul densității de repartiție (3.65) conduce la o observație care s-a dovedit a fi extrem de importantă în aplicațiile practice ale acestei repartiții, în cercetările statistice. Este vorba de faptul că densitatea de repartiție (3.65) nu depinde de dispersia σ^2 și aceasta face posibilă utilizarea repartiției Student în toate cazurile în care variabilele aleatoare independente $X_1, X_2, \ldots, X_n$ sînt normale, $N(0, \sigma)$, fără a avea vreo informație asupra dispersiei teoretice σ^2.

Fie $X_1, X_2, \ldots X_n$, n variabile aleatoare independente normale $N(\mathrm{m}, \sigma)$ și să însemnăm:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}. \tag{3.68}$$

Cum din teoria probabilităților se știe că o sumă de repartiții normale independente este tot o repartiție normală, iar o repartiție normală înmulțită cu o constantă dă naștere tot la o repartiție normală, $X_1, X_2, \ldots X_n$ fiind variabile independente $N(m, \sigma)$, rezultă din observația de mai sus că $\sum_{i=1}^{n} X_i$ are o repartiție normală și că $\frac{1}{n}\sum_{i=1}^{n} X_i$ are, de asemenea, o repartiție normală.

Se constată imediat că:

$$M(\bar{X}) = m$$

$$D^2(\bar{X}) = \frac{\sigma^2}{n}.$$

Să introducem notația:

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2. \tag{3.69}$$

Expresia:

$$u = \frac{\bar{x} - m}{\tilde{s}} \sqrt{n} \tag{3.70}$$

este de aceeaşi formă cu t, dat de formula (3.64) şi urmează o repartiţie Student cu $s = n - 1$ grade de libertate. Formula (3.70) are numeroase aplicaţii în practică.

Aplicaţia 21. (după H. Basler, *Grundbegriffe der Wahrscheinlichkeitsrechnung und statistischen Methodenlehre*, Phisica Verlag, Würzburg, Wien, 1970).

Se consideră greutatea medie X a n pachete de zahăr, umplute automat. S-a dovedit că X urmează o repartiţie normală $N(m, \sigma)$. În cazul în care s-au luat la întîmplare 7 pachete şi s-au obţinut valorile: $X_1 = 514$, $X_2 = 497$, $X_3 = 507$, $X_4 = 521$, $X_5 = 510$, $X_6 = 518$, $X_7 = 508$, se poate admite oare ipoteza că media teoretică a variabilei noastre aleatoare X este egală cu 500 ($m = 500$)?

Soluţie. Dacă $m = 500$, înseamnă că variabila aleatoare

$$u = \frac{\bar{X} - m}{\tilde{s}} \sqrt{n},$$

în care $\tilde{s}$ este dat de (3.69), urmează o repartiţie Student, cu 6 grade de libertate.

În acest caz, valorile experimentale luate de $\bar{X}$ şi $\tilde{s}$ sînt:

$$\bar{X} = \frac{517 + 497 + 507 + 521 + 510 + 518 + 508}{7} = 510{,}7$$

$$s^2 = \frac{1}{6}[(514 - 510{,}7)^2 + (497 - 510{,}7)^2 + \ldots + (508 - 510{,}7)^2 = 63{,}2$$

$$\tilde{s} = \sqrt{63{,}2} = 7{,}9. \tag{3.71}$$

Valoarea experimentală a lui u este:

$$\frac{510{,}7 - 500}{7{,}9} \sqrt{7} = 3{,}6. \tag{3.72}$$

Să notăm prin $u_{95\%}$ valoarea pentru care avem:

$$P[u \leqslant u_{95\%}] = 0{,}95.$$

Din tabelele existente pentru probabilităţile repartiţiei Student [13] găsim că pentru 6 grade de libertate $u_{95\%} = 2{,}45$ şi prin urmare:

$$P[u > u_{95\%}] = 0{,}05$$

$$P[u > 2{,}45] = 0{,}05. \tag{3.73}$$

Din (3.72) rezultă însă că am obţinut, pe baza unui singur eşantion, valoarea $3{,}6 > 2{,}45$.

Din (3.73) observăm că probabilitatea de a obţine o valoare mai mare decît 2,45 — ceea ce în cazul nostru s-a realizat — este 0,05.

[13] Vezi anexa 5.

Scanned with OKEN Scanner

Dacă într-adevăr $x = 500$, înseamnă că dintr-o singură probă s-a realizat un eveniment a cărui probabilitate de realizare este 0,05, adică o probabilitate foarte mică. Cum această eventualitate n-o considerăm plauzibilă, trebuie să înlăturăm ipoteza că greutatea medie, teoretică, a pachetelor este de 500 unități de măsură. În concluzie, cu o probabilitate egală cu 0,95 nu se admite ipoteza formulată.

Cu ajutorul repartiției Student putem rezolva și alte probleme care prezintă interes practic. [14]

Repartițiile „z" și „F" ale lui R. A. Fisher. R. A. Fisher a adus o contribuție remarcabilă la elaborarea metodologiei de cercetare a variațiilor, care au loc în domeniul fenomenelor naturale și social-economice. La baza acestei metodologii stă ideea comparării a doi indicatori tipici de variație: se compară fie două abateri medii pătratice, fie — ceea ce în majoritatea cazurilor este mai convenabil — două dispersii de sondaj. Se recurge la procedeul comparării, pentru a stabili dacă cele două dispersii pot fi considerate ca indicatori de estimație a dispersiei teoretice necunoscute a uneia și aceleiași populații originare, despre care se știe că se supune unei legi normale de repartiție. Problema care se pune — mereu aceeași în diferite cercetări de sondaj — este aceea de a ști dacă deosebirea între două dispersii de sondaj este sau nu este semnificativă.

1°. Să considerăm mai întîi procedeul comparării a două abateri medii pătratice introducînd în acest scop logaritmii naturali corespunzători acestora.

Indicatorul respectiv, notat cu z capătă forma:

$$z = \log_e s_1 - \log_e s_2, \tag{3.74}$$

unde s_1 și s_2 sînt, respectiv, abaterile medii pătratice în două eșantioane de volum n_1 și n_2.

Desigur z este o variabilă aleatoare.

Se poate demonstra că ea are următoarea densitate de repartiție [15]

$$f(z, f_1, f_2) = 2\left(\frac{f_1}{f_2}\right)^{\frac{f_1}{2}} \frac{\Gamma\left(\frac{f_1+f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)} e^{f_1 z}\left(1+\frac{f_1}{f_2}e^{2z}\right)^{-\frac{f_1+f_2}{2}}, \tag{3.75}$$

pentru $(-\infty < z < +\infty)$, în care f_1 și f_2 sînt, respectiv, numerele gradelor de libertate în două eșantioane, n_1 și n_2.

Repartiția caracterizată prin această densitate de repartiție se numește repartiție „z", a lui R.A. Fisher.

2° În majoritatea cercetărilor aplicative este mai comod să se facă comparație între dispersiile a două eșantioane, construindu-se în acest scop raportul lor. Indicatorul respectiv — notat cu F — capătă forma:

$$F = \frac{s_1^2}{s_2^2}, \tag{3.76}$$

unde s_1^2 și s_2^2 sînt estimații independente ale dispersiei teoretice σ^2 din populația originară, normal repartizată.

[14] Vezi aplicațiile date în [86] ca și în capitolul 5, consacrat teoriei sondajului.
[15] Vezi [60 p. 336].

Scanned with OKEN Scanner

Acest raport este o variabilă aleatoare a cărei densitate de repartiție este următoarea:

$$f(F)\ \frac{f_1^{\frac{f_1}{2}} f_2^{\frac{f_2}{2}} \Gamma\left(\frac{f_1+f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)(f_2+f_1F)^{\frac{f_1+f_2}{2}}} F^{\frac{f_1}{2}-1}. \tag{3.77}$$

Repartiția caracterizată prin această densitate de repartiție se numește repartiția F a lui R. A. Fisher. Este evident că valorile limită ale lui F sînt zero și $+\infty$

Forma curbei densității de repartiție a lui F, în limitele arătate, depinde de mărimea valorilor f_1 și f_2, — gradele de libertate cu care au fost calculate dispersiile de sondaj s_1^2 și s_2^2. De aceea, există curbe de repartiție ale lui F, simetrice față de axa absciselor (cînd $f_1 = f_2$) și asimetrice (cînd $f_1 \neq f_2$). Cota-parte procentuală a densității de repartiție a lui F, aflată la stînga unei mărimi date a lui F, funcția de repartiție F ($F < F_0$) poate fi găsită în tabele special întocmite în acest scop [16].

De pildă, unui $F < 0{,}4348$ îi corespunde 0,5% din aria delimitată de curba de repartiție $f(F)$ și axa absciselor; unui $F < 0{,}9833$ îi corespunde 50% din arie, iar unui $F < 1{,}9839$ îi corespunde 99,5%.

Date fiind definițiile lui z (3.74) și F (3.76) rezultă că între aceste mărimi există o strînsă legătură.

Se observă cu ușurință că între ele sînt valabile următoarele relații:

$$F = e^{2z} \quad \text{și} \quad z = \frac{1}{2}\log_e F. \tag{3.78}$$

3.5 REPARTIȚII MULTIDIMENSIONALE*

3.5.1 Generalități. Conceptul de sistem de variabile aleatoare

În paragrafele precedente ne-am ocupat de variabile aleatoare unidimensionale și repartițiile lor teoretice. În aplicațiile practice ale teoriei probabilităților întîlnim însă, adeseori, probleme în al căror context rezultatul unui experiment poate și trebuie să fie descris cu două sau mai multe variabile aleatoare, formînd astfel un *sistem* de asemenea variabile. Vom conveni să notăm un sistem de mai multe variabile aleatoare $(X, Y, Z, \ldots, W)$.

Proprietățile unui sistem de variabile aleatoare nu-și găsesc expresia, în mod exhaustiv, doar în proprietățile variabilelor aleatoare componente ale acestuia, ci și într-o serie de alte proprietăți, reflectînd *relațiile reciproce* dintre aceste variabile, dependențele lor lăuntrice.

În studiul sistemelor de variabile aleatoare, deosebit de utilă se dovedește a fi interpretarea lor geometrică. De pildă, un sistem de două variabile aleatoare (X, Y) poate fi reprezentat printr-un *punct aleator* pe un plan cu coordonatele X și Y. Analog se poate vorbi de un sistem de n variabile aleatoare, ca de „un punct aleator în spațial n-dimensional". De asemenea,

[16] Vezi anexa 6.

*Expunere după [130].

Scanned with OKEN Scanner

se dovedește utilă și interpretarea *vectorială* a unui sistem de variabile aleatoare, introducîndu-se în acest scop conceptul de *vector aleator*. De pildă, sistemul de două variabile aleatoare poate fi interpretat ca un vector aleator în planul XOY, ale cărui componente, pe axe, reprezintă variabilele aleatoare X și Y.

Teoria vectorilor aleatori — în prezent puternic închegată — sintetizează în categorii abstracte de mare interes practic atît legile de repartiție a probabilităților care caracterizează *complet* diferite sisteme, cît și numeroase valori tipice ale variabilelor aleatoare, în contextul cărora apar, totodată, repartiții condiționate, ca și *valori tipice condiționate*.

3.5.2 Repartiția sistemului de n variabile aleatoare

Analog variabilelor aleatoare unidimensionale, un sistem de n variabile aleatoare poate fi complet definit prin *legea sa de repartiție*. Această lege poate fi ilustrată fie prin funcția de repartiție n-dimensională, fie prin densitatea de repartiție de același tip.

Prin funcția de repartiție a unui sistem de n variabile aleatoare $(X_1, X_2, ..., X_n)$ *se înțelege probabilitatea realizării simultane a unui număr de n inegalități de tipul* $X_i < x_i$:

$$F(x_1, x_2, ..., x_n) = P[(X_1 < x_1)(X_2 < x_2), ..., (X_n < x_n)]. \tag{3.79}$$

Prin densitatea de repartiție a unui sistem de n variabile aleatoare continue se înțelege derivata mixtă de ordinul n a funcției de repartiție $F(x_1, x_2, ..., x_n)$:

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_2, x_1, ... x_n)}{\partial x_1 \partial x_2 ... \partial x_n} \tag{3.80}$$

Considerînd legea de repartiție a sistemului de variabile aleatoare putem defini legile de repartiție a fiecărei variabile în parte integrată în sistem. Funcția de repartiție a fiecărei variabile luată izolat se obține egalînd cu ∞, restul argumentelor din funcția de repartiție a sistemului

$$F_1(x) = F(x_1, \infty, ..., \infty). \tag{3.81}$$

Densitatea de repartiție a unei variabile din sistem se obține integrînd densitatea de repartiție a sistemului în limite infinite, pe întregul domeniu al restului de argumente

$$f_1(x_1) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1 \; x_2, ..., x_n)\, dx_2 ... dx_n. \tag{3.82}$$

Subsisteme de variabile aleatoare și repartițiile lor. Din sistemul de variabile aleatoare $X_1, X_2, ..., X_n$ putem separa un subsistem $X_1, X_2, ..., X_k$. Funcția de repartiție și densitatea de repartiție a acestui subsistem sînt definite, respectiv, de relațiile (3.83) și (3.84) de mai jos:

$$F_{1, 2, ..., k}(x_1, x_2, ..., x_k) = F(x_1, x_2, ..., x_k, \infty, ..., \infty) \tag{3.83}$$

$$f_{1, 2, ... k}(x_1, x_2, ..., x_k) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n)\, dx_{k+1} ... dx_n. \tag{3.84}$$

Scanned with OKEN Scanner

Pentru un subsistem de variabile aleatoare putem defini și conceptul de *repartiție condiționată.*

Vom spune că subsistemul $(X_1, X_2, ..., X_k)$ are repartiție condiționată dacă legea de repartiție se calculează după ce restul de variabile din sistem $(X_{k+1}, X_{k+2}, ..., X_n)$ au luat valorile $x_{k+1}, x_{k+2}, ..., x_n$.

Densitatea de repartiție condiționată poate fi calculată pe baza următoarei formule:

$$f(x_1, x_2, ..., x_k/x_{k+1}, x_{k+2}, ..., x_n) = \frac{f(x_1, x_2, ..., x_n)}{f_{k+1}, f_{k+2}, ..., n\,(x_{k+1}, ..., x_n)}. \tag{3.85}$$

Vom da acum o formulă generală cu ajutorul căreia putem calcula probabilitatea oricăror evenimente, indiferent dacă ele se încadrează sau nu în schema clasică a cazurilor egal *posibile* și a celor *favorabile.*

În esență este vorba deci de probabilitatea ca punctul aleator $(X_1, X_2, ..., X_n)$ să cadă în limitele spațiului D, n-dimensional:

$$P(X_1, X_2, ..., X_n \subset D) = \int p ... \int f(X_1\, X_2, ..., X_n)\, dx_1\, dx_2 ... dx_n. \tag{3.86}$$

De pildă, dacă evenimentul A nu poate fi încadrat în schema clasică a cazurilor, atunci acest eveniment se încadrează în schema variabilelor aleatoare — în majoritatea cazurilor, variabilele aleatoare continue — și în acest caz realizarea lui constă în faptul că sistemul de variabile $(X_1, X_2, ..., X_n)$ se va găsi în limitele domeniului D.

Aplicația 22. Un corp străin căzut pe parbrizul unui autoturism aflat în mers este în măsură să-l spargă dacă: a) mărimea unghiului θ cu care corpul străin ciocnește parbrizul se găsește între anumite limite (θ_1, θ_2); b) corpul are o greutate de cel puțin q_0 gr; c) viteza relativă a corpului străin la ciocnirea sa de parbrizul autoturismului nu este mai mică de v_m m/secundă. Ne punem problema de a calcula probabilitatea p ca un corp oarecare, căzut pe parbrizul mașinii, să producă efectul de a-l sparge. Este evident că viteza ciocnirii v, greutatea corpului q și unghiul de ciocnire θ formează un sistem de variabile aleatoare cu densitatea de repartiție $f(v, q, \theta)$.

Soluție. Integrînd densitatea de repartiție $f(v, q, \theta)$ în spațiul cu trei dimensiuni, corespunzător evenimentului de spargere a parbrizului, obținem probabilitatea căutată;

$$p = \int_{v_m}^{v_{max}} \int_{q_0}^{q_{max}} \int_{\theta_1}^{\theta_2} f(v, q, \theta) d_v\, d_q\, d_\theta.$$

În această expresie v_{max} și q_{max} simbolizează, respectiv, viteza și greutatea maximă a corpului străin.

3.5.3 Valori tipice ale unui sistem de variabile aleatoare

Adeseori, în practică, se ivesc situații (volumul limitat al informației experimentale, dificultatea calculelor etc.), cînd definirea completă a unui sistem de variabile prin legea sa de repartiție nu este posibilă și nici nu se justifică

Scanned with OKEN Scanner

în raport cu exigențele limitate ale exactității. Pe de altă parte, uneori ne este cunoscut tipul de repartiție (tipul repartiției normale, de pildă) care se aplică în cazul concret al problemei pe care o cercetăm. În toate aceste cazuri, descrierea sistemului de variabile o putem face cu ajutorul unui număr minim de valori tipice.

În acest sens putem considera:

1°. n valori medii, m_1, m_2, ..., m_n, caracterizînd mărimile medii ale fiecărei variabile;

2°. n dispersii, $D^2(X_1)$, $D^2(X_2)$, ..., $D^2(X_n)$, caracterizînd împrăștierea fiecărei variabile;

3°. n(n — 1) *momente de corelație*, definite astfel:

$$K_{ij} = M[(X_i - m_i) \cdot (X_j - m_j)], \tag{3.87}$$

caracterizînd corelațiile pe variabile-perechi ale întregului sistem.

Este evident că dispersia unei variabile aleatoare din sistem este un caz particular al momentului de corelație

$$D^2(X_i) = K_{ii} = M[(X_i - m_i)^2]. \tag{3.88}$$

Este deosebit de comod să ne folosim de o *matrice* a tuturor momentelor de corelație

$$\begin{bmatrix} K_{11} & K_{12} \dots & K_{1n} \\ K_{21} & K_{22} \dots & K_{2n} \\ \dots\dots\dots\dots & & \\ K_{n1} & K_{n2} \dots & K_{nn} \end{bmatrix}$$

pe care o vom numi *matricea de corelație* a sistemului de variabile aleatoare $(X_1, X_2, \dots, X_n)$.

Din modul în care am definit momentul de corelație rezultă că $K_{ij} = K_{ji}$; aceasta înseamnă că *elementele matricei de corelație, dispuse simetric față de diagonala ei principală, sînt egale.* De aceea, matricea de corelație poate fi prezentată și astfel:

$$\begin{bmatrix} K_{11} & K_{12} \dots & K_{1n} \\ & K_{22} \dots & K_{2n} \\ & \dots & \\ & \dots & K_{nn} \end{bmatrix}$$

Se observă că pe diagonala principală a matricei de corelație sînt dispuse dispersiile variabilelor aleatoare $X_1, X_2, \dots, X_n$.

În cazul în care variabilele aleatoare din sistem sînt *necorelate*, atunci toate elementele matricei de corelație sînt egale cu zero, afară de cele de pe diagonala principală. Avem deci, în acest caz, o matrice *diagonală* de formă:

$$\| K_{ij} \| = \begin{bmatrix} \sigma_1^2 & 0 & 0 \dots 0 \\ & \sigma_2^2 & 0 \dots 0 \\ & & \dots \\ & & \sigma_n^2 \end{bmatrix}.$$

Scanned with OKEN Scanner

Raţionamentele noastre cu privire la caracterul corelaţiilor din sistem le putem face mai concise, folosind o *matrice de corelaţie normată* $\| \rho_{ij} \|$, alcătuită nu din momente de corelaţie, ci din coeficienţii de corelaţie.

$$\rho_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j}, \tag{3.89}$$

în care

$$\sigma_i = \sqrt{D^2(X_i)}; \quad \sigma_j = \sqrt{D_2(X_j)}.$$

Matricea de corelaţie normată are forma:

$$\| \rho_{ij} \| = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1n} \\ & 1 & \rho_{23} & \dots & \rho_{2n} \\ & & 1 & \dots & \rho_{3n} \\ & & & \dots & \\ & & & \dots & 1 \end{bmatrix}$$

În mod firesc, elementele de pe diagonala principală sînt egale cu unitatea.

Vom preciza acum şi conceptul de vectori aleatori *necorelaţi*. Doi vectori aleatori X şi Y sînt *necorelaţi* dacă fiecare din elementele vectorului X nu este corelat cu fiecare element al vectorului Y.

$$K x_i y_i = M[(X_i - m_x)(Y_i - m_j] = 0, \quad (i, j = 1, 2, \dots, n).$$

3.5.4 Repartiţie multidimensională normală. Repartiţie condiţionată

Din ansamblul legilor de repartiţie, care caracterizează sisteme de două sau mai multe variabile aleatoare, un interes deosebit îl prezintă legea normală de repartiţie care se întîlneşte frecvent în cercetările aplicative.

În cazul general, densitatea de repartiţie, care caracterizează un sistem cu un număr oarecare de variabile aleatoare, are următoarea expresie:

$$f(x_1, x_2, \dots, x_n) = \frac{\sqrt{|C|}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}(x_i - m_{x_i})^2}, \tag{3.90}$$

în care: $|C|$ este determinantul matricei C, iar $C = \|C_{ij}\|$ este matricea *inversă* a matricei de corelaţie K, adică dacă matricea de corelaţie este:

$$K = \| K_{ij} \|$$

atunci

$$C_{ij} = (-1)^{i+j} \frac{M_{ij}}{|K|}.$$

În ultima expresie, $|K|$ este determinantul matricei de corelaţie, iar M_{ij} — minorul acestui determinant obţinut prin eliminarea rîndului i şi a coloanei j. Vom observa că

$$|C| = \frac{1}{|K|}.$$

Scanned with OKEN Scanner

Din expresia generală (3.90) rezultă toate expresiile legii normale pentru oricare n.

În cazul particular, cînd $n = 2$, matricea de corelaţie a două variabile aleatoare X şi Y este următoarea:

$$K = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y r \\ \sigma_x\sigma_y r & \sigma_y^2 \end{bmatrix},$$

r fiind coeficientul de corelaţie. Rezultă imediat şi determinantul acestei matrice şi al matricei C:

$$|K| = \sigma_x^2\sigma_y^2(1 - r)^2; \quad |C| = \frac{1}{\sigma_x^2\sigma_y^2(1 - r^2)}.$$

Matricea C este următoarea:

$$C = \begin{bmatrix} \frac{1}{\sigma_x^2(1 - r^2)} & \frac{-r}{\sigma_x\sigma_y(1 - r^2)} \\ \frac{-r}{\sigma_x\sigma_y(1 - r^2)} & \frac{1}{\sigma_y^2(1 - r^2)} \end{bmatrix}$$

Dacă introducem determinantul $|C|$ şi elementele lui în expresia generală (3.90) obţinem densitatea de repartiţie care caracterizează complet un vector aleator bidimensional normal,

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x - m_x)^2}{\sigma_x^2} - \frac{2r(x - m_x)(y - m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]} \tag{3.91}$$

Această lege depinde de cinci parametri: m_x, m_y, σ_x, σ_y şi r. Sensul acestor parametri este evident: m_x şi m_y sînt valorile medii (centrul împrăştierii) ale variabilelor X şi Y; σ_x, σ_y sînt abaterile medii pătratice respective, iar r — coeficientul de corelaţie.

Vom face acum o observaţie de importanţă deosebită:

Pentru un sistem de variabile aleatoare care se supune legii normale de repartiţie, din faptul că variabilele aleatoare sînt *necorelate* ($r = 0$) decurge totodată şi *independenţa* lor. Conceptele de variabile „necorelate“ şi variabile „independente“ — pentru cazul legii normale de repartiţie — sînt echivalente.

Cînd $r \neq 0$, variabilele aleatoare (X, Y) sînt dependente.

Pentru cazul repartiţiei normale bidimensionale vom defini şi conceptul de *repartiţie condiţionată.*

Să notăm prin $P(x_i/y_j)$ probabilitatea evenimentului $(X = x_i)$, condiţionată de evenimentul $(Y = y_j)$.

Prin repartiţia lui X, condiţionată de $Y = y_j$, se înţelege repartiţia:

$$\begin{pmatrix} x_1 & x_2 \ldots & x_m \\ p(x_1/y_j) & p(x_2/y_j) \ldots & p(x_m/y_j) \end{pmatrix}.$$

Analog putem defini şi repartiţia lui Y, condiţionată de $X = x_i$. Dacă nu este cunoscută repartiţia bidimensională, putem calcula probabilităţile $p(x_i/y_j)$ şi $p(y_j/x_i)$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$ pe baza formulei generale (3.3)

Scanned with OKEN Scanner

Astfel putem scrie:

$$p(x_i/y_j) = \frac{P(X = x_i;\ \ Y = y_j)}{P(Y = Y_j)} = \frac{p_{ij}}{p_{\cdot j}}$$

$$p(y_j/x_i) = \frac{p_{ij}}{p_{i\cdot}},$$

în care: $p_{\cdot j}$ și $p_{i\cdot}$ sînt *probabilități marginale.*

Repartițiile condiționate le putem caracteriza prin *valori tipice condiționate.*

De pildă, pentru variabile aleatoare discrete, media lui Y, condiționată de $X = x$, este dată de relația:

$$M(Y/X = x) = \sum_{j=1}^{m} y_j p(y_j/x)$$

sau de o relație mai generală — valabilă și pentru variabile continue — pe care o dăm fără demonstrație:

$$M(Y/X = x) = m_y + r\frac{\sigma_y}{\sigma_x}(x - m_x). \tag{3.92}$$

Totodată abaterea medie pătratică condiționată este definită de următoarea relație generală:

$$\sigma_{x/y} = \sigma_y\sqrt{1 - x^2}. \tag{3.93}$$

Formulele (3.92) și (3.93) pun în evidență faptul că în cazul unei legi de repartiție a lui Y, condiționată de valoarea $X = x$, de această valoare depinde numai valoarea medie și nicidecum dispersia.

Dependența (3. 92) poate fi reprezentată în planul xOy, însemnînd media condiționată $m_{y/x}$ pe axa ordonatelor. Se obține o dreaptă care se numește *dreapta de regresie* a lui Y asupra lui X. Analog, dreapta:

$$M(X/Y = y) = m_x + r\frac{\sigma_x}{\sigma_y}(y - m_y) \tag{3.94}$$

reprezintă regresia lui X asupra lui Y. Liniile de regresie coincid dacă există o relație funcțională liniară între X și Y. Dacă X și Y sînt variabile independente, liniile de regresie sînt paralele cu axele de coordonate.

Aplicația 23. Să considerăm următoarea repartiție bidimensională discretă:

y_i \ x_i	$x_1 = 1$	$x_2 = 2$	$x_3 = 4$	$x_5 = 5$	$p_{\cdot j}$
$y_1 = 3$	0,15	0,06	0,25	0,04	0,50
$y_2 = 4$	0,30	0,10	0,03	0,07	0,50
$p_{i\cdot}$	0,45	0,16	0,28	0,11	1,00

Scanned with OKEN Scanner

Să se determine — pe baza acestei repartiții — valoarea medie condiționată a lui Y, după ce a avut loc evenimentul $X = x_1$. Observăm că probabilitatea marginală $p_1.$ este dată de relația:

$$P_1. = 0{,}15 + 0{,}30 = 0{,}45.$$

Pe această bază putem calcula și repartiția lui Y, condiționată de $X = x_1$. Întrucît avem:

$$P(Y_1/x_1) = \frac{p_{11}}{p_1.} = \frac{0{,}15}{0{,}45} = \frac{1}{3}$$

$$P(Y_2/x_1) = \frac{p_{21}}{p_1.} = \frac{0{,}30}{0{,}45} = \frac{2}{3},$$

repartiția căutată este următoarea:

$$\begin{pmatrix} 3 & 4 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix};$$

pe baza ei calculăm media lui Y condiționată de $X = x_1$:

$$M(Y/X = x_1) = 3 \cdot \frac{1}{3} + 4 \cdot \frac{2}{3} = \frac{11}{3}.$$

3.6 LEGEA NUMERELOR MARI. TEOREMA LIMITĂ CENTRALĂ

3.6.1 Legea numerelor mari

După cum am văzut, în natură și în viața social-economică există numeroase fenomene de masă care se bucură de o anumită regularitate, stabilitate a frecvențelor și că conceptul de probabilitate este strîns legat de asemenea stabilități, constatate statistic.

Vom preciza acum că orice stabilitate — sintetizînd, în esență, o valoare medie — prezintă conținutul material al *legităților statistice*; el este pus în evidență cu atît mai puternic cu cît numărul elementelor componente ale fenomenului studiat este mai mare. Este vorba de acțiunea unei anumite legi obiective, care poartă numele de *lege a numerelor mari*.

În legătură cu sensul care se atribuie legii numerelor mari în teoria probabilităților și în alte discipline științifice sînt de făcut cîteva precizări.

În primul rînd, legii numerelor mari i se atribuie un sens larg: într-o *mare mulțime* de evenimente întîmplătoare, rezultatul lor *mediu* — practic vorbind — nu mai este întîmplător și de aceea poate fi studiat riguros din punct de vedere științific. Într-adevăr, particularitatea variabilelor aleatoare de a se comporta — în anumite condiții — ca și cum nu ar fi întîmplătoare, face posibilă *previziunea* evoluției acestora cu o deplină determinare. În aceasta își găsește expresia un aspect interesant al determinismului materialist-dialectic, și anume *determinismul statistic*. Un sens larg a fost atribuit legii

Scanned with OKEN Scanner

numerelor mari și de clasicii marxism-leninismului. Marx și Lenin, de pildă, au constatat și analizat acțiunea legii numerelor mari mai ales în corelație cu manifestările legii valorii [17]. Ma mult, din cercetările lui Engels [18] cu privire la natura ratei profitului și a legilor economice în general rezultă în mod evident că legea numerelor mari este o *formă de manifestare* a oricărei legi economice. În acest context este de reținut următorul aspect al legii numerelor mari: a) această lege ne indică doar *forma* specifică de manifestare a legităților de masă, dacă asemenea legități există; b) legea numerelor mari nu poate nici să creeze nivelul mediu al unui fenomen sau altul, nici să-l mențină sau să-l schimbe. Dacă, de pildă, a scăzut sau a crscut valoarea unei mărfi, este limpede că nu în acțiunea legii numerelor mari trebuie căutată cauza acestui fenomen, ci în însăși procesul de producție, în factorii determinanți, *esențiali*, ai acestui proces.

În al doilea rînd, legii numerelor mari i se atribuie un sens mai restrîns, constituit în cadrul teoriei probabilităților și al aplicațiilor acesteia. Este vorba de un grup de rezultate științifice extrem de importante, fiecare în parte stabilind, în anumite condiții, *convergența „în probabilitate" a valorilor medii ale unor experimentări de masă către niște constante bine determinate și care nu mai sînt întîmplătoare.*

Vom enunța în continuare, fără demonstrații, cîteva din aceste rezultate, avînd numeroase aplicații practice.

Inegalitatea lui Cebîșev. Adevărul exprimat prin această inegalitate este general, în sensul că-l întîlnim în întregul domeniu al variabilelor aleatoare, indiferent de tipul lor, discret sau continuu.

Fie X, o variabilă aleatoare. Ne propunem să estimăm probabilitatea evenimentului constînd în aceea că abaterea variabilei aleatoare X nu depășește — în valoare absolută — un număr pozitiv dat ε. Inegalitatea (3.95) ne indică răspunsul la această problemă:

$$P[\,|X - M(X)| < \varepsilon] \geqslant 1 - \frac{D^2(x)}{\varepsilon^2}. \tag{3.95}$$

De obicei, ε se ia în unități egale σ, adică, $\varepsilon = k\sigma$; în acest caz inegalitatea (3.95) ia următoarea formă:

$$P[\,|X - M(X)| < k\sigma] \geqslant 1 - \frac{1}{k^2}. \tag{3.96}$$

Pentru $k = 3$, de pildă, obținem:

$$P\,[\,|X - M\,(X)| < 3\sigma] \geqslant 1 - \frac{1}{9} = 0{,}888.$$

Deci, cu o probabilitate egală sau mai mare decît 0,888, putem afirma că abaterile în valoare absolută ale variabilei aleatoare X vor fi cuprinse în intervalul:

$$M\,(X) - 3\sigma, \quad M(X) + 3.$$

[17] Vezi [66, p. 67] și [78, p. 780].
[18] Vezi [43].

Scanned with OKEN Scanner

Teorema lui Cebîşev. Să considerăm un şir de variabile aleatoare independente, X_1, X_2, ..., X_n, ale căror dispersii sînt mărginite — mai mici decît o constantă. Teorema lui Cebîşev o putem formula astfel:

$$P\left[\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n}\right| < \varepsilon\right] \to 1. \qquad (3.97)$$

Această teoremă afirmă că oricare ar fi numărul pozitiv ε, probabilitatea inegalităţii (3.97) tinde către 1, dacă numărul n al variabilelor aleatoare tinde către infinit.

Adeseori în practică se iveşte cazul cînd variabilele $X_i (i = 1,2, \ldots, n)$ au aceeaşi medie a. Pentru acest caz particular, teorema lui Cebîşev afirmă că probabilitatea inegalităţii

$$\left|\frac{\Sigma(X_i)}{n} - a\right| < \varepsilon \qquad (3.98)$$

tinde către 1, dacă numărul n al variabilelor aleatoare tinde către infinit.

În esenţă, teorema lui Cebîşev pune în evidenţă faptul că între comportarea fiecărei variabile aleatoare în parte şi cea a mediei lor aritmetice există o mare deosebire: nu este posibil să prevedem ce valoare va lua fiecare variabilă aleatoare; însă, cu o probabilitate mare, putem prevedea ce valoare va lua *media* aritmetică a acestora. Putem trage de aici următoarea concluzie importantă.

Media aritmetică a unui număr suficient de mare de variabile aleatoare independente, avînd dispersii mărginite, îşi pierde caracterul de variabilă aleatoare.

Teorema lui Cebîşev confirmă teza filozofică materialist-dialectică cu privire la categoriile duale de necesitate — întîmplare, ca fundament ontologic al legităţilor statistice şi, implicit, al legii numerelor mari.

Vom vedea în capitolele consacrate metodei sondajului că această teoremă o fundamentează din punct de vedere teoretic.

Teorema lui Bernoulli. Această teoremă este de fapt un caz particular al teoremei lui Cebîşev.

Dacă se fac n experimente independente, în fiecare experiment, probabilitatea evenimentului A fiind p, atunci — cu o probabilitate oricît de apropiată de 1 — abaterea frecvenţei relative de apariţie a evenimentului A faţă de p va fi, în valoare absolută, oricît de mică, cu condiţia ca numărul n al experimentelor să fie suficient de mare.

Cu alte cuvinte, dacă ε este un număr pozitiv arbitrar de mic şi dacă notăm cu f_n numărul de apariţii ale evenimentului A, în cele n experimente, atunci:

$$\lim_{n\to\infty} P\left[\left|\frac{f_n}{n} - p\right| < \varepsilon\right] = 1, \quad (\varepsilon > 0). \qquad (3.99)$$

După cum se vede, teorema lui Bernoulli explică stabilitatea frecvenţelor în cazul unui număr mare de experimente, reiterate în aceleaşi condiţii.

Scanned with OKEN Scanner

3.6.2 Teorema limită centrală

În teoria probabilităţilor s-a formulat o teoremă celebră, cunoscută sub denumirea de *teorema limită centrală*. Acest rezultat s-a obţinut după îndelungi cercetări asupra erorilor de măsurare.

În esenţă, prin această teoremă se afirmă următorul adevăr:

Dacă o variabilă aleatoare X este suma unui număr mare de variabile aleatoare independente [19] *şi fiecare variabilă are o pondere mică în sumă, atunci legea de repartiţie a variabilei aleatoare X este foarte apropiată de o lege normală.*

A.M. Leapunov a formulat în felul următor această teoremă:

Se consideră un şir de variabile aleatoare independente $X_1, X_2, \ldots, X_n$; se introduc notaţiile:

$$a_k = M(X_k); \quad \sigma_k^2 = D^2(X_k); \quad \rho_k^3 = M(|X - m_k|^3), \quad 1 \leqslant k \leqslant n.$$

$$\sigma_{(n)}^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$$

$$\rho_{(n)}^3 = \rho_1^3 + \rho_2^3 + \ldots + \rho_n^3.$$

Dacă:

$$\lim_{n \to \infty} \frac{\rho_{(n)}}{\sigma_{(n)}} = 0$$

funcţia de repartiţie a variabilei:

$$\frac{X_1 + X_2 + \ldots + X_n - (a_1 + a_2 + \ldots + a_n)}{\sigma_{(n)}}$$

tinde către funcţia $\Phi(x)$ a lui Laplace, cînd $n \to \infty$. Teorema limită centrală domină întreaga teorie a repartiţiei probabilităţilor diferitelor erori de măsurare. Pe baza ei, oamenii de ştiinţă au ajuns la concluzia că funcţia de repartiţie normală poate fi luată drept *model teoretic* pentru cercetarea probabilistică a majorităţii fenomenelor aleatoare, din natură şi societate.

Grupul teoremelor definind probabilistic legea numerelor mari, alături de teorema limită centrală, pun în evidenţă eficienţa acestor rezultate. Aplicate în diverse domenii, aceste rezultate — fundamentale în teoria probabilităţilor — constituie puntea de legătură între această teorie şi fenomenele reale.

[19] Independenţa nu este însă o condiţie esenţială.

Scanned with OKEN Scanner

CAPITOLUL 4
Repartiții empirice

4.1. CONSTRUIREA REPARTIȚIILOR EMPIRICE. REPREZENTARE GRAFICĂ

Repartițiile empirice sînt expresia grupării unităților populației observate în funcție de caracteristici calitative sau cantitative. Prin alcătuirea lor se caută să se extragă, din datele culese, primele informații despre modul de structurare a populației și despre corelațiile ei structurale. Pentru a fi eficiente și comode, repartițiile se stabilesc după o serie de reguli, unele generale, cu aplicație la toate tipurile de repartiție, altele specifice fiecărui tip în parte.

Mai întîi, orice repartiție se construiește potrivit unei idei directoare sau unui interes de cunoaștere, practic ori teoretic. Aceasta implică operații de selecționare în optica interesului reținut: în unele cazuri, repartiția se formează în raport cu o singură caracteristică, iar în alte cazuri, în raport cu două sau mai multe caracteristici; uneori se adoptă o gradație naturală a modalităților caracteristicii, alteori, una conceptuală sau convențională [1]. Numărul modalităților unei caracteristici poate de asemenea să varieze conform informației dorite. Modalitățile aceleiași caracteristici trebuie însă să fie totdeauna incompatibile — să se excludă reciproc și totodată exhaustive — să acopere toate cazurile particulare. Respectarea regulii din urmă apără de erori în clasarea unităților și în interpretarea repartiției.

Procesul de construire a repartițiilor empirice este un proces de sintetizare a observațiilor primare, care atrage după sine sacrificarea unei anumite cantități de informație. Există astfel riscul ca structurile reale să fie deformate și, implicit, repartițiile construite să devină inutilizabile. Aici regula constă în păstrarea la maximum a integrității informațiilor cuprinse în documentația de bază. Formarea de grupe omogene — o altă regulă elementară — presupune tocmai o condensare a datelor cu măsură și, în mod corespunzător, minimizarea pierderilor de informație.

Regulile de construire a repartițiilor empirice diferă apoi după cum caracteristicile cercetate sînt calitative sau cantitative.

[1] Modalitățile sînt diferitele situații posibile ale unei caracteristici. După cum ele se referă la calitatea sau, dimpotrivă, la intensitatea caracteristicii, modalitățile se numesc calitative sau cantitative. Fiecare caracteristică studiată poate prezenta două sau mai multe modalități.

Scanned with OKEN Scanner

4.1.1 Cazul unei caracteristici calitative

Cînd caracteristica este calitativă, repartiţia are forma generală:

Modalitatea caracteristicii	a	b	...	i	...	n
Frecvenţa fiecărei modalităţi	f_a	f_b	...	f_i	...	f_n

Caracteristica observată poate apărea cu modalităţi absolut definite (sexul: masculin — feminin; starea civilă: necăsătorit — căsătorit — văduv — divorţat; etc.) şi atunci repartiţia se construieşte fără dificultate. Dar, adesea, modalităţile caracteristicii nu pot fi circumscrise decît graţie admiterii anumitor convenţii, avînd deci un caracter mai mult sau mai puţin subiectiv (categorii socioprofesionale, ramuri ale economiei naţionale etc.). În asemenea circumstanţe sînt necesare eforturi pentru a defini diversele modalităţi în aşa fel încît să se preîntîmpine posibilităţile de interpretare personală.

Elementul prin care repartiţia intră în judecăţile statistice este *frecvenţa*. Sub forma prelucrată de frecvenţe, datele statistice sînt „semne sau manifestări ale unei structuri interioare ascunse, susceptibilă de a fi definită prin probabilităţi, a căror măsură este sau prezumată şi verificabilă prin frecvenţe, sau dedusă din aceste frecvenţe“ [2]. Frecvenţele caracterizează importanţa părţilor componente ale populaţiei şi asocierile caracteristicilor calitative sînt precizate prin comparaţii între frecvenţe.

Trăsătura distinctivă a repartiţiei după o caracteristică de tip calitativ este *simetria claselor*, în sensul că toate clasele se consideră egale între ele. Diferenţa între o clasă de unităţi sau alta constă în faptul de a poseda sau a nu poseda o modalitate prescrisă [3].

Exemplul 1. Repartiţia populaţiei României pe medii, în anii 1948 şi 1974.

Tabelul 4.1

Mediul	1948*		1974**	
	Numărul locuitorilor	%	Numărul locuitorilor	%
Total	15 872 624	100,0	21 028 841	100,0
Urban	3 713 139	23,4	8 978 917	42,7
Rural	12 159 485	76,6	12 049 924	57,3

* Recensămînt (25.I)
** Estimare (1.VII).
Sursa: *Anuarul statistic al Republicii Socialiste România 1975*, p. 9.

Avem de-a face cu o repartiţie clar exhaustivă şi dihotomică, adică formată doar din două clase. Între cele două clase nu există însă o delimitare naturală: localităţile de tip urban, respectiv de tip rural, sînt specificate ca atare prin legislaţia şi reglementările din fiecare ţară, de unde şi distincţia *definiţie naţională* şi *definiţie internaţională* a populaţiei urbane.

[2] O. ONICESCU, vezi lucrarea citată la p. 32.
[3] Pentru a arăta că nu există nici o deosebire de grad între clasele care formează repartiţia (adică $a = b = \ldots = i = \ldots = n$), aceasta a primit şi denumirea de *homogradă*.

Scanned with OKEN Scanner

Ca să fie mai evocatoare, repartiția se poate transpune grafic sub forma unui cerc divizat în sectoare proporționale cu ponderile (fig. 4.1.).

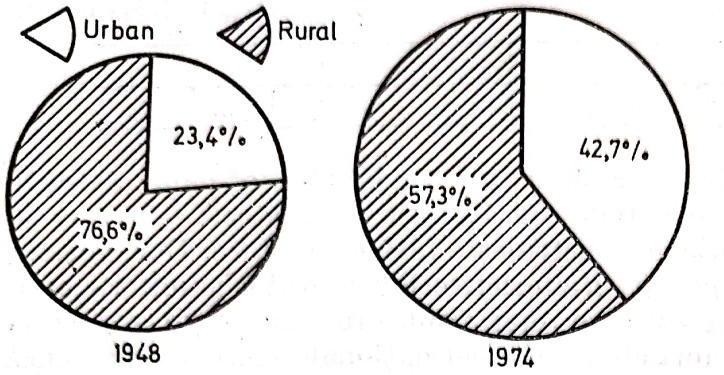


Fig. 4.1

4.1.2 Cazul unei caracteristici cantitative

Dacă $x_1, x_2, \ldots, x_i, \ldots, x_n$ sînt valorile observate pentru o caracteristică sau variabilă (discretă sau continuă) X și $f_{(x_1)}, f_{(x_2)}, \ldots, f_{(x_i)}, \ldots, f_{(x_n)}$ frecvențele corespunzătoare, repartiția variabilei X se scrie:

$$X \begin{pmatrix} x_1, & x_2, \ldots, x_i, \ldots, x_n \\ f_{(x_1)}, & f_{(x_2)}, \ldots, f_{(x_i)}, \ldots, f_{(x_n)} \end{pmatrix}.$$

Bineînțeles, valorile luate de variabila statistică se înșiruie în ordinea mărimii, crescătoare și cîteodată descrescătoare [4].

În cazul unei variabile discrete și cu un număr restrîns de valori, fiecărei valori particulare a acesteia îi corespunde o anumită frecvență (corespondență univocă).

Exemplul 2. Repartiția născuților vii, după rang, în R.S. România, în anul 1974.

Tabelul 4.2

Rangul născutului viu	Numărul născuților vii		Freevența absolută cumulată		Frecvența relativă cumulată	
	nr. abs	%	ascendentă	descendentă	ascendentă	descendentă
1	159 327	37,3	159 327	427 672	37,3	100,0
2	120 299	28,1	279 626	268 345	65,4	62,7
3	68 247	16,0	347 873	148 046	81,4	34,6
4	42 015	9,8	389 888	79 799	91,2	18,6
5	16 935	3,9	406 823	37 784	95,1	8,8
6	8 298	1,9	415 121	20 849	97,0	4,9
7	5 009	1,2	420 130	12 551	98,2	3,0
8 și peste	7 542	1,8	427 672	7 542	100,0	1,8

Sursa: *Anuarul statistic al Republicii Socialiste România 1975*, p. 38.

[4] Unitățile înregistrate se diferențiază în raport cu mărimea caracteristicii și de aceea repartiția rezultată se întîlnește sub denumirea *heterogradă* (terminologie introdusă de C. Charlier, statistician suedez, vezi [47, p. 24]).

Scanned with OKEN Scanner

O astfel de repartiţie se reprezintă grafic prin diagrama în batoane sau în bare (fig. 4.2.): pe axa absciselor se notează valorile variabilei (rangul născutului viu) şi se ridică pentru fiecare valoare o ordonată de lungime proporţională cu numărul de unităţi observate (numărul născuţilor vii). Pentru

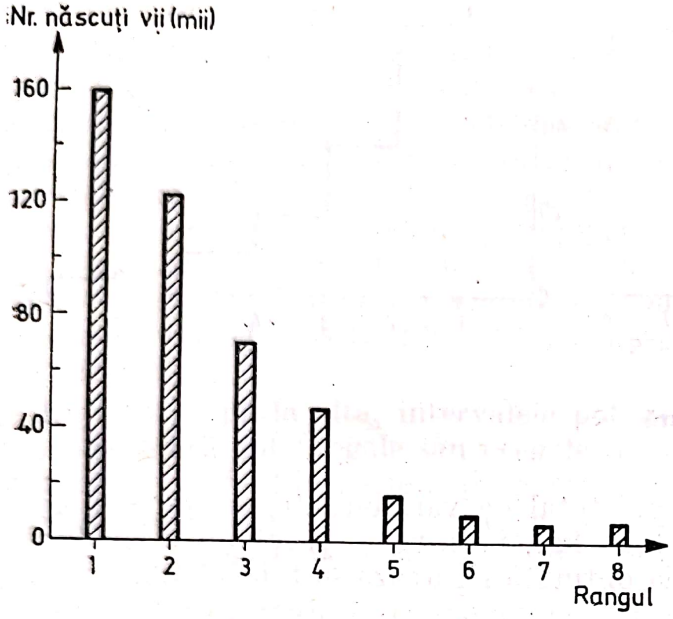


Fig. 4.2

compararea a două sau mai multe repartiţii cu frecvenţe absolute inegale, se recurge la repartizarea frecvenţelor relative, modificîndu-se în consecinţă scara de-a lungul axei verticale.

Alături de frecvenţele simple, absolute şi relative, s-au introdus frecvenţele cumulate, în două variante: însumarea succesivă a frecvenţelor simple cu începere de la prima grupă — frecvenţe cumulate ascendente şi cu începere de la ultima grupă — frecvenţe cumulate descendente (în sensul citirii normale a repartiţiei). Ele poartă denumirea de funcţie de repartiţie empirică (experimentală), simbolizată prin $F_{(x_i)}$ şi exprimă numărul sau proporţia unităţilor statistice care posedă o valoare a variabilei mai mică sau egală cu x_i, în prima variantă, şi mai mare sau egală cu x_i, în a doua variantă. În exemplul luat, frecvenţele cumulate, înscrise bunăoară la grupa a 3-a, arată că, printre femeile care au dat naştere la copii vii în anul 1974, existau 81,4% (347 873) în acel moment mame a 1—2 copii şi 34,6% (148 046), mame a 3 şi mai mulţi copii. Diagramele cumulative (fig. 4.3 şi 4.4) au alura unei scări; înălţimea fiecărei trepte este proporţională cu o frecvenţă simplă.

De regulă, variabilele observate apar cu o gamă largă de valori şi în acest caz se aplică repartiţia pe o scară cu intervale.

Exemplul 3. Pe baza datelor despre producţia şi suprafaţa cultivată cu grîu şi secară de fiecare întreprindere agricolă de stat, dintr-un număr de 244 i.a.s., s-au calculat producţiile medii la hectar şi s-a aflat că se eşalonează

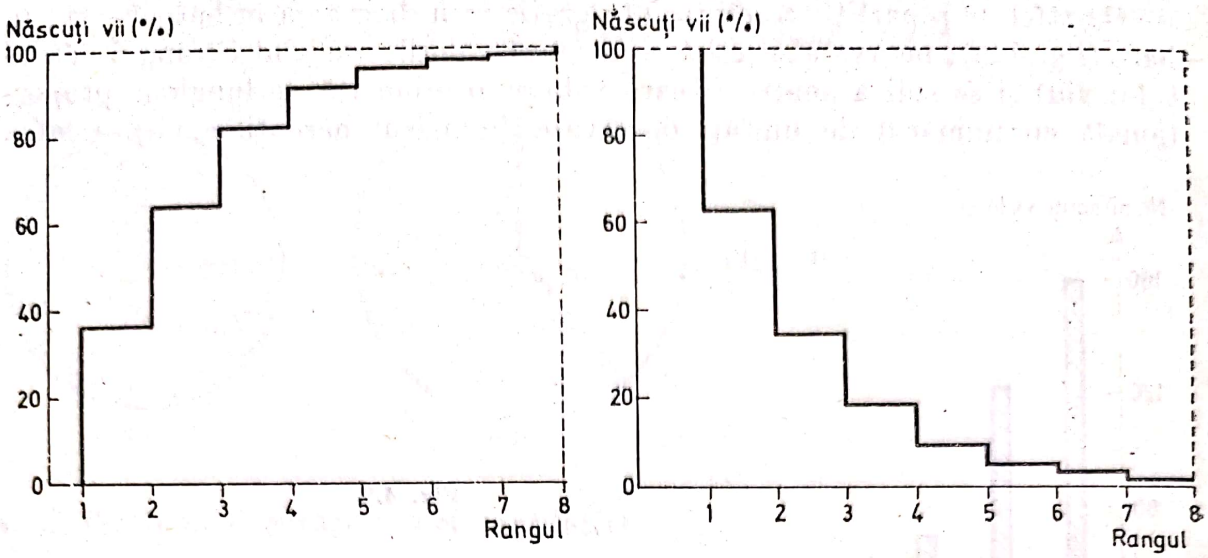


Fig. 4.3
Fig. 4.4 } **Diagrame cumulative în scară**

între 1 422 kg și 3 758 kg. Cîmpul total de variabilitate s-a împărțit în 12 intervale egale cu dimensiunea (200 kg), i.a.s.-urile grupîndu-se în funcție de intervalele stabilite. Repartiția obținută figurează în tabelul 4.3.

Tabelul 4.3

Producția medie la hectar (kg)*	Număr de i.a.s.	Frecvența cumulată	
		ascendentă	descendentă
1400—1600	4	4	244
1600—1800	21	25	240
1800—2000	27	52	219
2000—2200	33	85	192
2200—2400	39	124	159
2400—2600	41	165	120
2600—2800	32	197	79
2800—3000	24	221	47
3000—3200	10	231	23
3200—3400	7	238	13
3400—3600	4	242	6
3600—3800	2	244	2

* Medii pe trei ani (1967—1969).
Sursa: [90, p. 46].

Caracteristica „producția medie la hectar" variază de o manieră cvasi-continuă și dacă s-ar opera pe o scară cu valori distincte, repartiția ar fi extrem de întinsă. Cu aplicație la colecția de unități tratată, chiar rotunjirile în chintale (numere întregi) ar fi condus la o repartiție compusă din 25 grupe — complicată și neexpresivă. Ca urmare, s-a impus întrunirea „producțiilor medii la hectar" pe o scară de repartiție cu intervale.

Problema nouă și de primă importanță, generată deci de repartiția pe intervale, privește determinarea grupelor de valori ce alcătuiesc scara cla-

Scanned with OKEN Scanner

sificării unităților observate. Trebuie avut în vedere faptul că fixarea intervalelor de grupare include decuparea mai mult sau mai puțin arbitrară a cîmpului de variație al caracteristicii, cu implicațiile ei. Astfel, o scară greșit aleasă poate schimba complet aspectul repartiției. Și aici, experiența cercetătorului și buna cunoaștere a domeniului investigat au menirea să joace rolul esențial.

Multă grijă comportă decizia în alegerea, printre un mare număr de soluții posibile, a mărimii intervalelor de grupare. Desigur, nu se poate da o indicație universală în ceea ce privește mărimea optimă a intervalelor; numai o analiză prealabilă, pentru fiecare caz în parte, este în stare să ofere un răspuns. Se recomandă doar ca intervalele să nu fie prea largi, întrucît ar produce o pierdere mare de informație și, implicit, ar micșora precizia rezultatelor și nici prea înguste, căci, deși ar păstra o informație mai deplină, n-ar estompa iregularitățile accidentale, în schimb ar complica fără real folos calculele. Un echilibru trebuie să fie găsit între „o precizie amăgitoare și o informație suficientă"[5].

De la o grupă la alta, intervalele pot avea o amplitudine constantă sau variabilă, adică pot fi egale sau neegale. Este de preferat repartiția pe o scară cu intervale egale, fiindcă numai așa frecvențele diferitelor grupe sînt comparabile între ele și potrivite pentru calculele ulterioare. Nu întotdeauna însă împărțirea pe intervale egale este cea adecvată.

Exemplul 4. Repartiția întreprinderilor din industria de subordonare republicană după numărul muncitorilor, în R.S. România, în anii 1950 și 1974.

Tabelul 4.4

Număr de muncitori	1950		1974	
	Număr de întreprinderi	%	Număr de întreprinderi	%
Total	1 136	100,0	1 251	100,0
Pînă la 200	535	47,1	164	13,1
201—500	295	26,0	190	15,2
501—1000	170	15,0	278	22,2
1001—2000	95	8,4	325	26,0
2001—3000	21	1,8	153	12,2
3001—5000	13	1,1	96	7,7
Peste 5000	7	0,6	45	3,6

Sursa: *Anuarul statistic al Republicii Socialiste România 1957*, p. 77 și *1974*, p. 86.

Intervalele marginale sînt deschise; intervalele intermediare au amplitudinile: 300, 500, 1000, 1000, 2000 ... muncitori. Adoptarea de intervale variabile ca mărime, după poziția lor în scară, a fost dictată de faptul că între întreprinderi există mari diferențe sub raportul numărului de muncitori.

[5] Numărul optim de intervale, admis de unii autori, se situează între 10 și 15, iar de alții, între 15 și 25 [134, p. 96]. S-a imaginat și o formulă empirică pentru a calcula un număr de clase optimal K, atunci cînd n reprezintă numărul total de observații [87]:

$$K = 1 + 3{,}3 \log_{10} n.$$

Scanned with OKEN Scanner

Scara aleasă permite precizarea frecvențelor cu care se înscriu întreprinderile mici, mijlocii și mari în ansamblul întreprinderilor și, prin comparații în timp, cunoașterea dimensiunilor procesului de concentrare industrială.

În principiu, o scară cu intervale de mărime egală presupune o populație perfect omogenă în ceea ce privește compoziția; în caz contrar, intervalele se iau de mărimi diferite, pentru ca diferențierile calitative în structura populației să nu fie eclipsate.

Limitele intervalelor de grupare trebuie să fie explicit indicate, ca să se suprime orice risc de eroare. De exemplu, forma utilizată la repartiția din tabelul 4.4 este explicită, dar forma de prezentare din tabelul 4.3, în care valoarea limitei superioare a unui interval apare și ca valoare a limitei inferioare a intervalului următor, poate da naștere la ambiguități. Într-adevăr, valorile observate ce s-ar găsi la frontiera grupelor ar putea fi repartizate în chip diferit (cum ar fi: i.a.s. -ul cu producția medie la hectar de 1 600 kg s-ar putea repartiza și în prima și cea de-a doua grupă). Neajunsul se înlătură printr-o convenție prealabilă, care, în general, admite cuprinderea în interval a limitei sale inferioare și excluderea limitei superioare.

Cît privește reprezentarea grafică a repartițiilor pe o scară cu intervale, forma frecvent utilizată este *histograma* (fig. 4.5.) [6]. Pentru fiecare grupă a repartiției se construiește un dreptunghi avînd baza proporțională cu intervalul de grupă (valorile caracteristicii) și înălțimea proporțională cu frecvența (absolută sau relativă); suma suprafețelor dreptunghiurilor reprezintă suma tuturor frecvențelor. Uneori se substituie histogramei în scară *poligonul frecvențelor* (fig. 4.5.), unind succesiv, prin segmente de dreaptă, punctele mediilor ale treptelor (un punct are ca abscisă valoarea medie a intervalului și ca ordonată frecvența). Suprafața cuprinsă între poligonul frecvențelor și axa absciselor este echivalentă cu aria histogramei.

Cînd intervalele sînt suficient de mici și populația relativ mare, profilul de linie poligonală tinde către o curbă continuă numită *curba frecvențelor*. Această curbă ideală nu are sens decît dacă însăși variabila variază continuu.

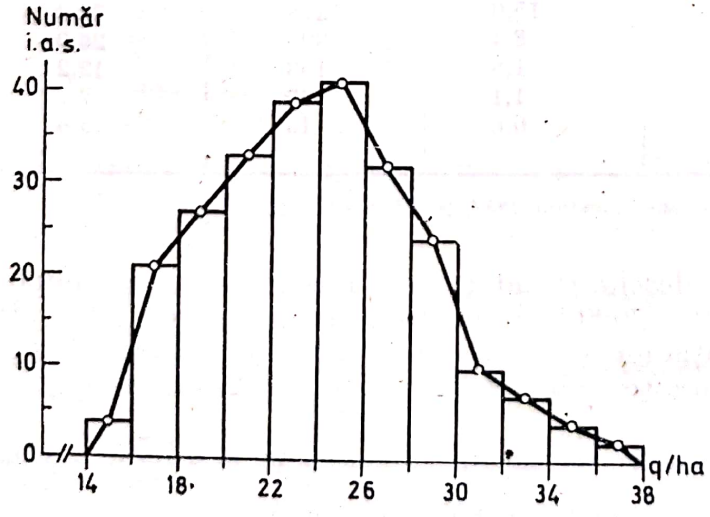


Fig. 4.5

[6] Termen introdus de K. Pearson.

Scanned with OKEN Scanner

Repartițiile cu intervale neegale cer o anumită transformare înainte de a fi reprezentate grafic. Să reluăm repartiția din tabelul 4.4 (anul 1974).

Tabelul 4.5

Număr de muncitori	Număr de întreprinderi	Mărimea intervalului	Număr de intervale unitare	Frecvența pe interval unitar
0	1	2	3	4
Pînă la 200	164	200	1,0	164,0
201—500	190	300	1,5	126,6
501—1000	278	500	2,5	111,2
1001—2000	325	1 000	5,0	65,0
2001—3000	153	1 000	5,0	30,6
3001—5000	96	2 000	10,0	9,6
Peste 5000	45	2 000	10,0	4,5

Propunîndu-ne să reprezentăm repartiția printr-o histogramă cu menținerea proporționalității între suprafețele dreptunghiurilor și frecvențe, am ales un interval unitar (primul interval = 200), a cărei amplitudine să fie divizor comun tuturor celorlalte intervale. Construim histograma (fig. 4.6) ridicînd dreptunghiuri care au bazele proporționale cu intervalele de grupă, dar înălțimile proporționale cu frecvențele pe interval unitar (col. 4).

Repartiția sub formă cumulativă (tabelul 4.3) se reprezintă grafic după principiul aplicat în reprezentarea unei repartiții simple: la extremitatea fiecărei clase se ridică o ordonată proporțională cu frecvența cumulată pînă la această limită și unindu-se două cîte două punctele înscrise rezultă *poligonul frecvențelor cumulate* (fig. 4.7). Este ușor de scos în evidență pe linia poligonală valorile caracteristicii asociate diferitelor clase sau, la limită, diferitelor unități cumulate în ordinea crescătoare sau descrescătoare a caracteristicii [7].

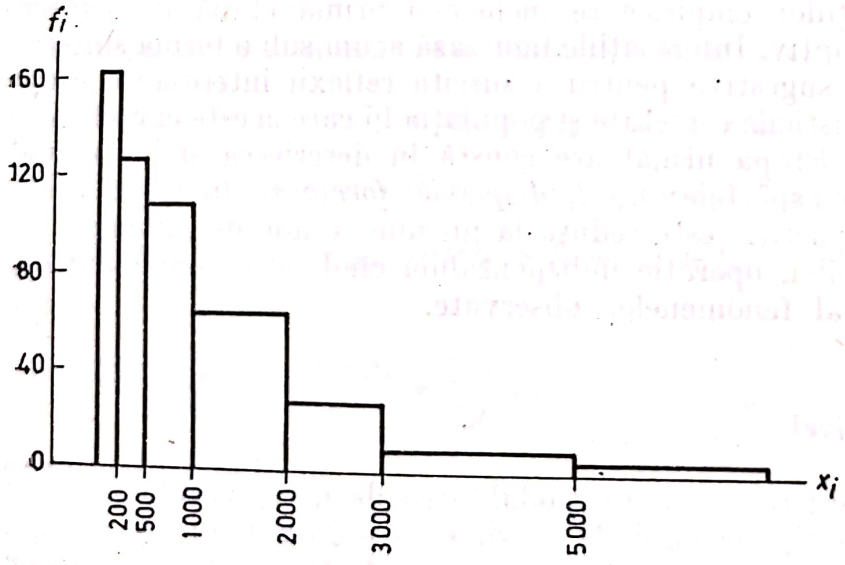


Fig. 4.6

[7] Curba frecvențelor cumulate, obișnuit de forma unui *S* alungit, se mai numește *ogivă*, nume dat în arhitectură curbei de aceeași formă.

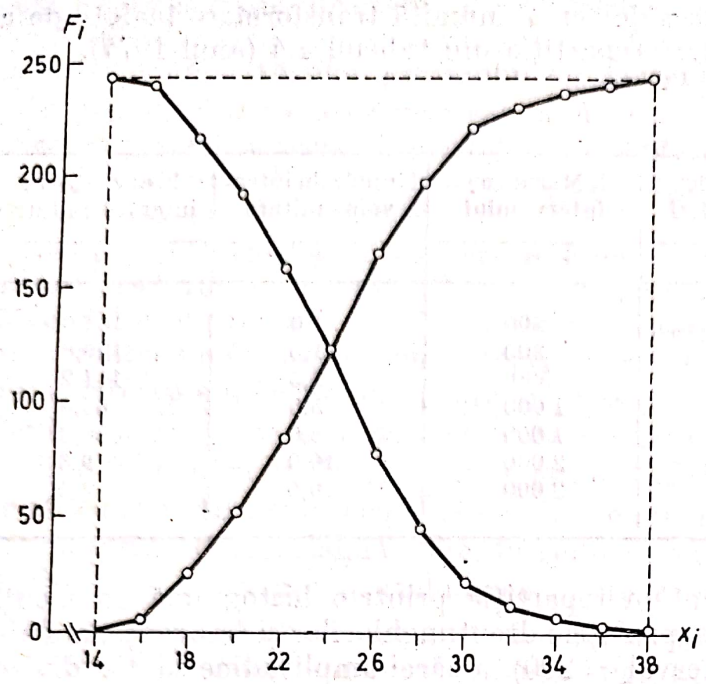


Fig.4.7. Curbele frecvenţelor cumulate—ascendente şi descendente.

4.2 DESCRIEREA NUMERICĂ A REPARTIŢIILOR EMPIRICE

Cu alcătuirea repartiţiilor empirice se încheie o primă etapă în demersul statistic de ordin descriptiv. Informaţiile figurează acum sub o formă sintetică, suficient de simplă şi sugestivă pentru a suscita reflexii interesante asupra relaţiilor între caracteristicile cercetate şi populaţia în care aceste caracteristici sînt variat împărţite. Etapa următoare constă în descrierea şi individualizarea repartiţiilor sub aspectele: *nivel*, *dispersie*, *formă* şi, în unele cazuri, *concentrare*. Fiecare repartiţie este redusă la un număr mic de *parametri* sau *valori tipice* ale variabilei, operaţie indispensabilă cînd dorim să pătrundem în studiul structural al fenomenelor observate.

4.2.1 Parametrii de nivel

În scop de primă descriere, se caută modalitatea de a reprezenta printr-un număr unic, o valoare-tip, ordinul de mărime al ansamblului de elemente ce compun repartiţia statistică. O atare valoare poate fi considerată mărimea variabilei corespunzătoare frecvenţei maxime sau mărimea variabilei situate la mijlocul şirului de valori aranjate în ordine crescătoare (sau descrescătoare) sau, în fine, mărimea medie a variabilei deduse într-un anumit fel din toate datele repartiţiei.

Media. Conceptul de medie este familiar. De obicei, cuvîntul „medie", fără nici o altă specificare, se întrebuinţează în accepţia de *medie aritmetică* —

Scanned with OKEN Scanner

medie fundamentală în teorie și comodă în practică. Toată lumea știe să calculeze media unei serii de date; natural, fiecare valoare trebuie să intervină în calcul de atîtea ori cît ea a fost constatată, adică de atîtea ori cîte unități prezintă frecvența respectivă. Se poate spune că media aritmetică este nivelul mediu ponderat al populației în raport cu variabila considerată.

Conform repartiției din tabelul 4.2, rangul mediu al născutului viu a fost în anul 1974 egal cu:

$$\frac{1}{427\,672}[(1\times159327) + (2\times120299) + (3\times68247) + (4\times42015) +$$

$$+ (5\times16935) + (6\times8298) + (7\times5009) + (8\times7542)] = 2{,}3.$$

În mărimea rangului mediu al născutului viu se reflectă sintetic nivelul fertilității. Grăitoare în ce privește diferențele de fertilitate, de exemplu pe categorii sociale, sînt rangurile medii ale născuților vii calculate pentru același an: 3 la țărani cooperatori, 2,1 la muncitori și 1,6 la funcționari (față de 2,3 la întreaga populație).

Relativ la media calculată se impun două observații:

a) în raport cu valorile concrete pe care le sintetizează, apare ca o mărime abstractă;

b) substituită fiecăreia din valorile repartiției, suma lor nu se schimbă.

Într-adevăr, ca valoare sintetică, media se obține printr-un proces de abstractizare; calculul mediei înseamnă abstragerea de la nivelurile individuale ale caracteristicii, considerînd-o egală ca mărime pentru fiecare unitate a populației. Astfel media „al 2,3-lea rang" este o mărime abstractă, care egalează născuții vii în raport cu rangul nașterii. Dealtfel, se poate ca valoarea medie să nu existe în realitate la nici una din unitățile observate, ca în exemplul luat [8].

Cea de-a doua remarcă, exprimată în relația matematică

$$x_1f_1 + x_2f_2 + \ldots + x_nf_n = \bar{x}f_1 + \bar{x}f_2 + \ldots + \bar{x}f_n,$$

ne conduce la formula de calcul a mediei, în cazul nostru medie aritmetică:

$$x_1f_1 + x_2f_2 + \ldots + x_nf_n = \bar{x}(f_1 + f_2 + \ldots + f_n),$$

de unde

$$\bar{x} = \frac{x_1f_1 + x_2f_2 + \ldots + x_nf_n}{f_1 + f_2 + \ldots + f_n} = \frac{\Sigma xf}{\Sigma f}, \tag{4.1}$$

care se mai poate scrie:

$$\bar{x} = x_1\frac{f_1}{\Sigma f} + x_2\frac{f_2}{\Sigma f} + \ldots + x_n\frac{f_n}{\Sigma f} = \Sigma x\frac{f}{\Sigma f}, \tag{4.2}$$

$\frac{f}{\Sigma f}$ fiind frecvențele relative.

[8] J. Bertillon a reliefat principala caracteristică a mediei — aceea de a fi o valoare abstractă: statisticianul cînd o calculează creează o mărime nouă [în *Cours élémentaire de statistique*, Paris, 1895].

Scanned with OKEN Scanner

Aplicată la repartiţia din tabelul 4.2.

$\bar{x} = (1 \times 0{,}373) + (2 \times 0{,}281) + (3 \times 0{,}16) + (4 \times 0{,}098) + (5 \times 0{,}039) + (6 \times 0{,}019) + (7 \times 0{,}012) + (8 \times 0{,}018) = 2{,}3.$

Modificarea proporţională a frecvenţelor nu modifică valoarea medie.

Calculul mediei unei repartiţii pe intervale include o ipoteză suplimentară cu privire la valoarea x-ilor: se presupune distribuirea uniformă a unităţilor în interiorul intervalului dat şi, în consecinţă, se admite că valorile x sînt reprezentate prin centrele intervalelor.

Să considerăm repartiţia din tabelul 4.3.

Tabelul 4.6

Centrul intervalului x	Frecvenţa f	xf	$\frac{x - 2700}{200}$	$\frac{x - 2700}{200} f$
1 500	4	6 000	−6	−24
1 700	21	35 700	−5	−105
1 900	27	51 300	−4	−108
2 100	33	69 300	−3	−99
2 300	39	89 700	−2	−78
2 500	41	102 500	−1	−41
2 700	32	86 400	0	0
2 900	24	69 600	1	24
3 100	10	31 000	2	20
3 300	7	23 100	3	21
3 500	4	14 000	4	16
3 700	2	7 400	5	10
Total	244	586 000	—	−364

Centrul intervalului se determină simplu ca semisumă a valorilor extreme.

Producţia medie la hectar este:

$$\bar{x} = \frac{586\,000}{244} = 2401{,}6 \text{ kg.}$$

Cum intervalele au aceeaşi amplitudine, calculele se pot prescurta, efectuînd o schimbare de variabilă: se înlocuieşte variabila x printr-o variabilă z, aşa încît

$$x = x_0 + d \cdot z;$$

x_0 şi d sînt constante.

Procedeul constă în aplicarea combinată a proprietăţilor mediei aritmetice care permit ca fiecare valoare a variabilei (x) să fie micşorată o dată prin scăderea unei constante ($x - x_0$) şi a doua oară prin împărţirea restului obţinut la o altă constantă $[(x - x_0) / d]$. Practic, calculele se fac pe variabila $z = (x - x_0) / d$, ale cărei valori depind de alegerea celor două constante x_0 şi d. Formula mediei aritmetice devine

$$\bar{x} = x_0 + d\bar{z}$$

şi desfăşurat $\bar{z}$

$$\bar{x} = x_0 + d \frac{\Sigma \left(\frac{x - x_0}{d} \right) f}{\Sigma f}. \qquad (4.3)$$

Scanned with OKEN Scanner

Mersul operațiilor este arătat în tabelul 4.6.

$$\bar{x} = 2700 + 200\,\frac{-364}{244} = 2700 - \frac{72\,800}{244} = \frac{658\,800 - 72\,800}{244} =$$

$$= \frac{586\,000}{244} = 2401{,}6 \text{ kg.}$$

Cantitatea cu care se scad termenii seriei (x_0), numită și „origine", poate fi oricare dintre termenii seriei și chiar o mărime din afara ei. Pentru a se evita însă abaterile prea mari și a se simplifica cît mai mult calculele, este indicată alegerea originii în așa fel încît să împartă scara repartiției în două părți aproximativ egale. Drept constantă d se alege mărimea intervalului de grupare, care dă, pentru noua variabilă z, numerele cele mai simple.

Procedeul schimbării variabilei ușurează calculele mai ales în cazul repartițiilor cu numere mari sau care comportă mai multe zecimale.

Evident, media calculată pentru o repartiție cu intervale exprimă, cu o oarecare aproximație, mărimea medie reală a variabilei, deoarece nu poate fi vorba de o corespondență perfectă între ipoteza de calcul admisă și realitate. Gradul de aproximație este în funcție de două elemente: uniformitatea repartiției unităților din grupe în interiorul intervalelor și dimensiunea intervalelor de grupare. Media va fi cu atît mai precisă cu cît repartiția unităților cuprinse în fiecare interval este mai uniformă și cu cît mărimea intervalelor este mai redusă. Repartițiile statistice cu o scară de intervale prea mici nu permit o evaluare satisfăcătoare a mediei. [9]

O formă particulară a mediei aritmetice este *media armonică.*

Exemplul 5. În scopul determinării unei norme de timp, s-a cronometrat, într-o uzină, o operație de bază în fabricarea unui produs. Datele înregistrate în cursul mai multor cronometrări succesive, repartizate în grupe de valori egale, sînt trecute în tabelul 4.7.

Tabelul 4.7

Consum specific de muncă (om-ore/operație) t	Număr de om-ore lucrate T	$\frac{1}{t}T$
0,80	56	70
0,85	204	240
0,90	288	320
0,95	380	400
1,00	430	430
1,05	315	300
1,10	143	130
1,15	92	80
1,20	36	30
Total	1944	2000

[9] O relație recomandată pentru stabilirea intervalului de grupare (d) este următoarea:

$$d = \frac{1}{100} \cdot 8(x_{max} - x_{min}),$$

în care x_{max} și x_{min} sînt valoarea cea mai mare și valoarea cea mai mică a seriei de valori, iar 8 este un factor empiric. Folosită în construirea repartiției din tabelul 4.3.

$$i = \frac{1}{100} \cdot 8\,(3758 - 1422) = 186{,}88 \text{ kg.}$$

Scanned with OKEN Scanner

Pentru a calcula consumul specific mediu de muncă potrivit datelor disponibile trebuie avută în vedere relaţia $T/t = Q$ (Q — număr de operaţii). Proprietatea determinantă a mediei are forma:

$$\frac{T_1}{t_1} + \frac{T_2}{t_2} + \ldots + \frac{T_n}{t_n} = \frac{T_1}{\bar{t}} + \frac{T_2}{\bar{t}} + \ldots + \frac{T_n}{\bar{t}}$$

sau

$$\frac{1}{t_1} T_1 + \frac{1}{t_2} T_2 + \ldots + \frac{1}{t_n} T_n = \frac{1}{\bar{t}} (T_1 + T_2 + \ldots + T_n),$$

de unde ecuaţia mediei

$$\bar{t} = \frac{T_1 + T_2 + \ldots + T_n}{\frac{1}{t_1} T_1 + \frac{1}{t_2} T_2 + \ldots + \frac{1}{t_n} T_n} = \frac{\Sigma T}{\Sigma \frac{1}{t} T};$$

deci

$$\bar{t} = \frac{1944}{2000} = 0{,}972 \text{ om-ore.}$$

Exprimată în formula generală

$$\bar{x} = \frac{f_1 + f_2 + \ldots + f_n}{\frac{1}{x_1} f_1 + \frac{1}{x_2} f_2 + \ldots + \frac{1}{x_n} f_n} = \frac{\Sigma f}{\Sigma \frac{1}{x} f}. \qquad (4.4)$$

Media armonică este mărimea inversă a mediei aritmetice calculată din mărimile inverse ale valorilor repartiţiei [10]. Utilizarea ei devine obligatorie în cazul relaţiilor de proporţionalitate inversă, ca în exemplul evocat. Într-adevăr, dacă privim variabila t derivată din T/Q, ne aflăm în prezenţa unei relaţii inverse: consumul specific de muncă (t), faţă de aceeaşi cantitate de cheltuieli de muncă (T joacă rol de pondere), creşte pe măsura descreşterii volumului de operaţii (Q) şi invers. În varianta în care ponderarea lui t s-ar face cu Q, existenţa unei proporţionalităţi directe — t faţă de aceeaşi mărime a lui Q creşte pe măsura creşterii lui T şi invers — justifică aplicarea mediei aritmetice. Adoptînd $T = tQ$ şi notaţiile din formula generală (x în loc de t şi f în loc de Q), se vede limpede că în funcţie de datele disponibile se calculează fie media aritmetică, fie media armonică şi se ajunge la aceeaşi valoare medie:

$$\bar{x} = \frac{\Sigma xf}{\Sigma \frac{1}{x} xf} = \frac{\Sigma xf}{\Sigma f}$$

De aici concluzia: alegerea mediei pretinde o atentă analiză a datelor problemei. [11]

[10] Se spune că media armonică a fost introdusă în teoria muzicii pentru compararea lungimii coardelor vibrante cu înălţimea sunetului lor. „Cum numărul vibraţiilor este *invers proporţional* cu lungimea coardei, nota medie, care corespunde numerelor de vibraţii date de două coarde de lungimi diferite, este dată de o coardă a cărei lungime este media armonică a lungimilor" [109, p. 217—218].

[11] Alte forme de medii sînt tratate în cap. 7.

Scanned with OKEN Scanner

Mediana. Reamintim că mediana este egală cu valoarea centrală a repartiției statistice avînd valorile variabilei așezate în ordine de mărime. Ea împarte colecția de unități observate în două grupe egale ca număr.

Pentru o repartiție de $2k + 1$ valori, mediana este valoarea ce are rangul $k + 1$. Dacă repartiția cuprinde $2k$ valori, mediana se situează între valorile de rangul k și $k + 1$ și, în general, s-a convenit să fie reprezentată de media aritmetică a celor două valori.

De fapt, într-o repartiție pe valori distincte, mediana apare determinată: în exemplul 2, rangul median al născutului viu este 2. Cînd datele sînt grupate pe intervale și pierd deci individualitatea lor, mediana se calculează prin procedeul interpolării liniare, bazat pe ipoteza repartiției uniforme (sau echirepartiție) a frecvențelor în intervalul median.

Exemplul 5. Să calculăm vîrsta mediană a soțului și vîrsta mediană a soției la încheierea căsătoriei, corespunzător repartițiilor pe anul 1974, în R. S. România.

Tabelul 4.8

Grupa de vîrstă (ani)	Căsătorii			
	soț		soție	
	Frecvența simplă	Frecvența cumulată	Frecvența simplă	Frecvența cumulată
Sub 20	8 592	8 592	65 694	65 694
20—24	92 856	101 448	76 437	142 131
25—29	44 816	146 264	15 976	158 107
30—34	12 516	158 780	5 629	163 736
35—39	5 933	164 713	3 686	167 422
40—44	3 159	167 872	2 555	169 977
45—49	2 217	170 089	1 985	171 962
50—54	1 496	171 585	1 513	173 475
55—59	906	172 491	744	174 219
60 și peste	3 005	175 496	1 277	175 496

Sursa: *Anuarul statistic al Republicii Socialiste România 1975*, p. 46—47.

Lectura frecvențelor cumulate ne arată că valorile de rangul $k = 87748$ și $k + 1 = 87749$ sînt cuprinse în intervalul al doilea „20—24 ani". Admițînd ipoteza echirepartiției, împărțim acest interval cu frecvența:

(a) $\frac{5}{92\,856}$ ani (b) $\frac{5}{76\,437}$ ani.

Pînă la valoarea de rangul $(87748 + 87749) / 2 = 87748{,}5$ revin intervalului median:

(a) 79156,5 frecvențe (87748,5 — 8592)

(b) 22054,5 frecvențe (87748,5 — 65694).

Partea din mărimea intervalului median adunată la limita sa inferioară [12] ne dă valoarea medianei (Me). Astfel:

(a) vîrsta mediană a soţului ... $19{,}5 \frac{5}{92\,856} \times 79156{,}5 = 23{,}76$ ani;

(b) vîrsta mediană a soţiei ... $19{,}5 \frac{5}{76\,437} \times 22054{,}5 = 20{,}94$ ani.

Formula de calcul a medianei poate fi scrisă astfel:

$$Me = x_k + d_k \frac{1/2(\Sigma f + 1) - F_{k-1}}{f_k}, \qquad (4.5)$$

unde

x_k este limita inferioară a intervalului median;
d_k — mărimea intervalului median;
f_k — frecvenţa simplă a intervalului median;
F_{k-1} — frecvenţa cumulată a intervalului imediat inferior intervalului median.
Σf — suma frecvenţelor repartiţiei.

Curbele frecvenţelor cumulate ascendente şi descendente permit o determinare grafică rapidă a medianei: mediana este determinată prin intersecţia acestor două curbe (fig. 4.8 şi 4.9.)

Modul. Modul sau dominanta [13] exprimă mărimea tipică a variabilei, aceea care corespunde frecvenţei celei mai mari. De vreme ce se repetă de cele mai multe ori, valoarea modală reprezintă şi valoarea cea mai „probabilă".

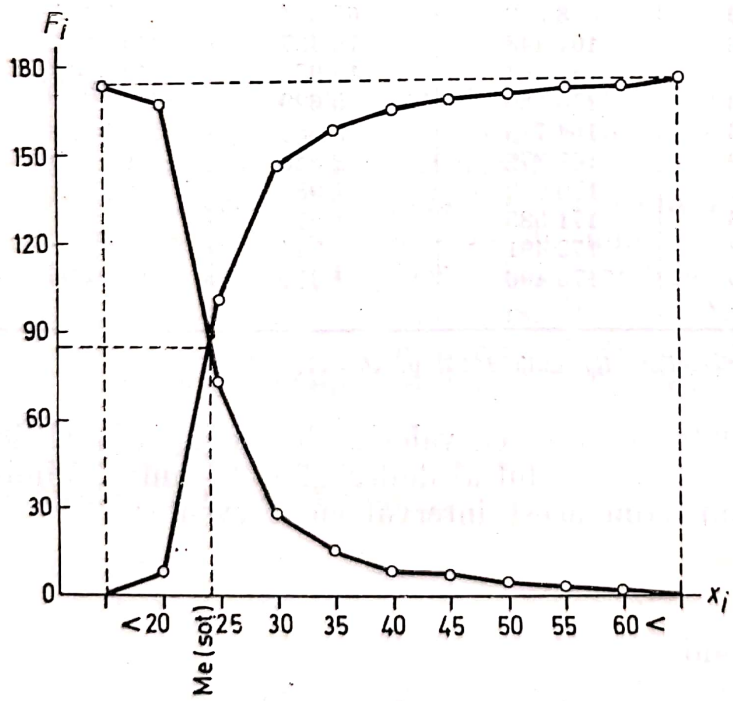


Fig. 4.8 Determinarea grafică a medianei

[12] Se decide, în general, că intervalele delimitate în forma întrebuinţată la repartiţia pe grupe de vîrstă nu încep în realitate de la valorile indicate în scară. De exemplu, intervalul vizat, 20—24, începe cu 19,5, pentru a se termina cu 24,5, această din urmă valoare marcînd limita inferioară a intervalului următor. Valoarea centrală, 22, este echidistantă faţă de cele două limite ale intervalului (respectiv 19,5 şi 24,5).

[13] Termenul englezesc *mode*, introdus de K. Pearson, a fost tradus în franţuzeşte prin acela de *dominante*.

Scanned with OKEN Scanner

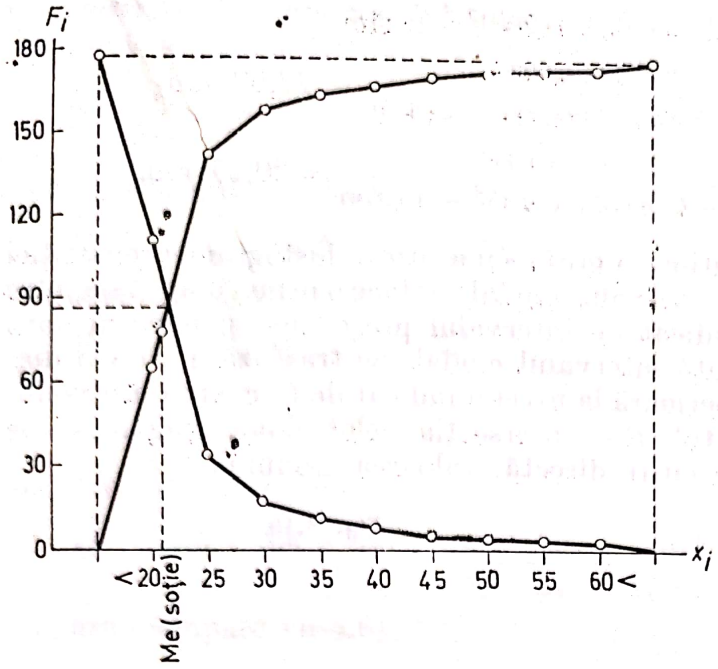


Fig. 4.9 Determinarea grafică a medianei

La repartițiile de valori individuale, determinarea modului este imediată, pe calea simplei examinări a frecvențelor. Astfel, rangul modal al născutului viu are mărimea 1 în repartiția din tabelul 4.2. Dacă repartiția îmbracă forma grupării pe intervale, este preferabil să vorbim mai degrabă de „interval modal" sau „interval dominant", ca, de exemplu, în cazul repartiției precedente (intervalul modal 20—24 ani, cu frecvența 92856 pentru soț și frecvența 76437 pentru soție). O estimare mai precisă a modului necesită anumite ipoteze și calcule.

În condiția unei repartiții simetrice sau atunci cînd intervalele imediat învecinate cu cel modal s-ar deosebi puțin unul de altul după frecvență sau greutate specifică, se poate lua, cu o aproximație satisfăcătoare, mijlocul intervalului modal ca valoare modală. Cînd repartițiile sînt de alt tip, modul se va găsi pe scara de variație a intervalului modal mai aproape sau mai departe de centrul acestui interval, în funcție de forma repartiției. Poziția modului în interiorul intervalului modal se determină în raport cu abaterea frecvenței intervalului premodal și a celui postmodal de la frecvența intervalului modal. Procedeul de calcul este un procedeu de interpolare, constînd în aplicarea formulei următoare:

$$Mo = x_k + d_k \frac{f_k - f_{k-1}}{(f_k - f_{k-1}) + (f_k - f_{k+1})}, \tag{4.6}$$

în care:

x_k este limita inferioară a intervalului modal;
d_k — mărimea intervalului modal;
f_k — frecvența intervalului modal;
f_{k-1} — frecvența intervalului premodal;
f_{k+1} — frecvența intervalului postmodal.

Scanned with OKEN Scanner

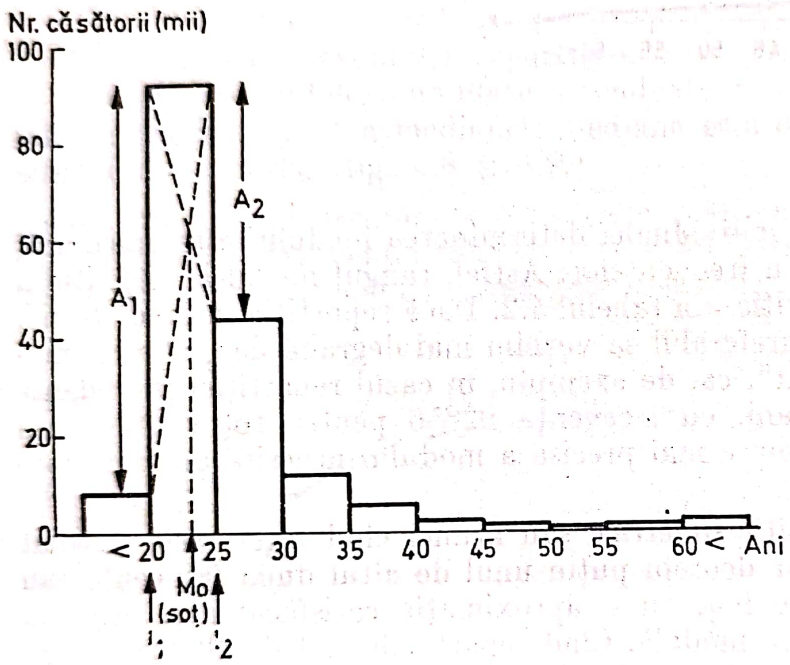
Vîrstele modale ale repartiţiilor din tabelul 4.8 sînt:

$$\text{— soţ ... } 19{,}5 + 5\,\frac{92\,856 - 8\,592}{(92\,856 - 8\,592) + (92\,856 - 44\,816)} = 22{,}69 \text{ ani,}$$

$$\text{— soţie ... } 19{,}5 + 5\,\frac{76\,437 - 65\,694}{(76\,437 - 65\,694) + (76\,437 - 15\,976)} = 20{,}25 \text{ ani.}$$

Valoarea modală poate fi estimată grafic cu ajutorul histogramei repartiţiei studiate (fig. 4.10. şi 4.11). În intervalul modal se duce o primă diagonală între limita inferioară, la nivelul indicat de intervalul precedent, şi punctul opus al dreptunghiului care reprezintă intervalul modal. Se trasează o a doua diagonală pornind de la limita superioară la nivelul indicat de frecvenţa intervalului următor. Proiectînd punctul P — intersecţia celor două diagonale, pe axa orizontală se obţine, prin citire directă, valoarea modului.

Fig. 4.10



Determinarea grafică a modului.

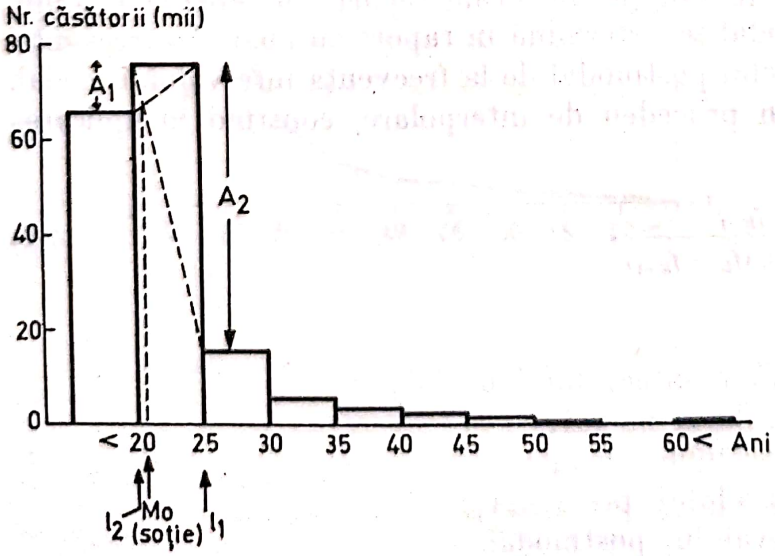


Fig. 4.11

Scanned with OKEN Scanner

Diagrama ilustrează metoda de interpolare a valorilor pentru determinarea modului. Folosind notaţiile pe grafic, scriem proporţia:

$$\frac{Mo - l_1}{l_2 - Mo} = \frac{\Delta_1}{\Delta_2},$$

care desfăşurată în felul următor

$$\Delta_2 Mo - \Delta_2 l_1 = \Delta_1 l_2 - \Delta_1 Mo,$$
$$\Delta_1 Mo + \Delta_2 Mo = \Delta_1 l_2 + \Delta_2 l_1,$$
$$Mo(\Delta_1 + \Delta_2) = \Delta_1 l_2 + \Delta_2 l_1$$

şi

$$l_2 = l_1 + d$$

ne conduce la formula modului

$$Mo = \frac{\Delta_1 l_1 + \Delta_1 d + \Delta_2 l_1}{\Delta_1 + \Delta_2} = \frac{\Delta_1 l_1 + \Delta_2 l_1}{\Delta_1 + \Delta_2} + \frac{\Delta_1 d}{\Delta_1 + \Delta_2} = l_1 + d\frac{\Delta_1}{\Delta_1 + \Delta_2},$$

în care (compară cu 4.6)

l_1 este x_k
$\Delta_1 = f_k - f_{k-1}$
$\Delta_2 = f_k - f_{k+1}$.

Un alt procedeu pentru aflarea modului, propus de K. Pearson, se sprijină pe relaţia care există între mod, mediană şi medie în repartiţiile moderat asimetrice. În orice repartiţie simetrică unimodală (un singur maximum), media aritmetică, mediana şi modul sînt identice; într-o repartiţie uşor asimetrică şi unimodală, se constată că mediana se găseşte între medie şi mod, distanţa sa faţă de mod fiind, aproximativ, dublul distanţei sale de la media aritmetică.

Prin urmare, cunoscîndu-se media aritmetică şi mediana, modul se deduce simplu din formula:

$$Mo = 3Me - 2\bar{x}. \qquad (4.7.)$$

Exemplul 6. Să considerăm repartiţia a 250 muncitori după retribuţia orară.

Tabelul 4.9

Retribuţia medie orară (lei)	Număr de muncitori	Număr cumulat
8–9	7	7
9–10	12	19
10–11	28	47
11–12	39	86
12–13	82	168
13–14	47	215
14–15	21	236
15–16	9	245
16–17	5	250

$\bar{x} = 12{,}41$ lei şi $Me = 12{,}48$ lei.

Scanned with OKEN Scanner

Modul, calculat după formula (4.6.):

$$Mo = 12 + 1 \frac{82 - 39}{(82 - 39) + (82 - 47)} = 12{,}55 \text{ lei};$$

iar după formula (4.7.):

$$Mo = (3 \times 12{,}48) - (2 \times 12{,}41) = 12{,}62 \text{ lei}.$$

Diferența dintre valorile obținute este nesemnificativă.

Proprietăți comparate. Condiții de utilizare. Parametrii de nivel [14] s-au impus în practica cercetării statistice datorită însușirii lor de a condensa substanța unei serii de date într-un singur număr. Ei sînt indispensabili atunci cînd dorim să comparăm două sau mai multe repartiții de frecvențe și să facem judecăți despre nivelul de dezvoltare al variabilei sau variabilelor reflectate.

Multă vreme, media, mediana și modul au reprezentat forma de statistică practicată obișnuit și de aceea s-a putut spune că „tehnica valorilor tipice" nu comportă decît „în mod cu totul excepțional puncte de întrebare științifice".

Cît privește întrebuințarea acestor parametri, este necesar să se cunoască proprietățile fiecăruia, pentru ca, ținînd seama de repartiția particulară tratată, să se aleagă parametrul cel mai semnificativ.

Statisticienii englezi Yule și Kendall au precizat o serie de condiții, care definesc un model ideal de medie [15], și anume:

a) să fie definită în mod obiectiv, independent de aprecierea cercetătorului;
b) să fie expresia tuturor termenilor repartiției;
c) să posede proprietăți simple și evidente, făcînd posibilă înțelegerea sensului ei general;
d) să poată fi calculată cu ușurință și rapiditate;
e) să fie afectată cît mai puțin cu putință de fluctuațiile de selecție;
f) să se preteze la calcule algebrice, mai ales cînd este vorba de compararea mai multor repartiții.

Dintre parametrii cunoscuți, *media aritmetică* se apropie cel mai mult de modelul imaginat, ceea ce și explică utilizarea sa curentă. Ea este o funcție de toate valorile observate (ca de altfel orice formă de medie: armonică, geometrică etc.), admite o teoremă de adițiune, se bucură de proprietăți matematice simple, permite calcule algebrice ulterioare, nu-i decît puțin sensibilă la fluctuațiile de selecție și este ușor de conceput.

Firește, dedusă din toți termenii repartiției, media aritmetică își pierde stabilitatea în prezența unor valori „aberante", anormal de mari sau anormal de mici. Acest „inconvenient" obligă la asigurarea premisei de omogenitate a datelor pentru care se calculează media. Nu trebuie să se uite că o medie reflectă cu atît mai precis proprietățile tipice ale populației cu cît mai omogenă este populația. Relativ la calculul mediilor („mijlociilor"), socotit „una din combinările cele mai obișnuite de statisticieni" — scria cu peste un secol în urmă I. Ionescu de la Brad — „... este foarte îmbelșugat în iluzii mai ales cînd extremele, cînd minimele și maximele sînt foarte depărtate. În acest caz

[14] Denumiri echivalente: parametrii tendinței centrale; valori tipice; caracteristici ale valorii centrale; măsuri (parametri) de poziție. De obicei, denumirea parametrii de poziție se atribuie numai medianei și modului.

[15] Prin medie sînt desemnați toți parametrii de nivel: medie, mediană, mod.

Scanned with OKEN Scanner

rezultatele sînt pe cît de străine, pe atîta şi de neadevărate. Mijlocia este o cantitate fictivă, dînd despre realitate o idee cu atît mai exactă, cu cît elementele din care se alcătuieşte sînt mai aproape între sine" [59, p. 7].

Prin urmare, dependenţa mediei aritmetice de toate valorile variabilei pune problema eliminării valorilor care s-ar părea excepţionale, accidentale, aberante [16].

Mediana, prin definiţie, nu ţine seamă de mărimea tuturor termenilor repartiţiei şi ea este deci mai puţin afectată de mărimea termenilor extremi. O unitate cu o valoare excepţională ca mărime influenţează totalul valorilor din care se calculează media aritmetică; aceeaşi unitate nu influenţează mediana decît prin rangul ei. [17]

Ca ilustrare a proprietăţii medianei de a avea o mai mare stabilitate decît media aritmetică, să reluăm repartiţia din tabelul 4.9, pentru care

$$\bar{x} = 12{,}41 \text{ lei}; \quad Me = 12,\ 48 \text{ lei}$$

şi să convenim la excluderea ultimei grupe „16—17 lei" cu frecvenţa 5; atunci:

$$\bar{x} = 12{,}32 \text{ lei}; \quad Me = 12{,}45 \text{ lei}.$$

În timp ce media aritmetică s-a micşorat cu 0,7%, mediana s-a micşorat cu numai 0,2%, adică în raport de 3,5 la 1.

Calculul medianei este uşor şi rapid, dar ea nu se pretează la operaţii algebrice. Mediana unei repartiţii globale nu poate fi dedusă din medianele repartiţiilor componente. Ea are însă o proprietate importantă: suma abaterilor tuturor valorilor repartiţiei în raport cu mediana, socotită în valoare absolută, este minimum. Această sumă este inferioară sumei obţinute măsurînd abaterile (în valoare absolută) de la oricare altă valoare, media în particular.

Un avantaj practic al medianei, deloc neglijabil, constă în faptul că poate fi calculată chiar dacă nu se cunosc în mod precis limitele extreme ale primului şi ultimului interval al repartiţiei considerate (repartiţie deschisă).

Cu toate insuficienţele ei, mediana — „media de poziţie" — este „un eşantion remarcabil al repartiţiei, remarcabil pentru că-i „la mijloc", utilizat într-un mare număr de analize statistice: studiul fertilităţii, mortalităţii, determinarea duratei mediei de viaţă etc., ca şi în probleme în care datele se referă la caracteristici calitative.

[16] În literatura statistică se recomandă un test foarte simplu pentru a descoperi „intrusul" (sau „intruşii") într-o serie de date statistice: se scade cea mai mică valoare a seriei din cea mai mare, mai întîi incluzînd valoarea suspectă, apoi eliminînd-o. Dacă rezultatul cu valoarea suspectă este cel puţin dublul rezultatului fără valoarea suspectă, atunci aceasta din urmă nu aparţine seriei. Un exemplu simplificat:

103; 115; *223*; 137; 143; 138; 119.

223 − 103 = 120 şi 143 − 103 = 40, o abatere de la simplu la triplu; deci 223 este o mărime străină.
sau

103; 115; *31*; 137; 143; 138; 119.

143 − 31 = 112 şi 143 − 103 = 40; 112: 40 = 2,8, deci 31 este o mărime străină. Se înţelege că testul singur nu poate constitui o justificare suficientă în eliminarea din serie a valorilor „suspecte" [Vezi: W. E. DUCKWORTH, *Méthodes statistiques de la recherche technologique*, Dunod, Paris, 1973, p. 14].

[17] Mediana este numită şi *medie de poziţie*, deoarece ea reţine doar locul, şi nu mărimea termenilor seriei.

Scanned with OKEN Scanner

Modul răspunde şi mai puţin condiţiilor lui Yule şi Kendall: nu se determină în funcţie de toate observaţiile, nu permite calcule algebrice, este foarte instabil.

Calculat pentru repartiţii cu un mare număr de termeni şi cu o tendinţă centrală netă, modul are totuşi o valoare descriptivă importantă: exprimă mărimea cea mai comună a variabilei, mărime a cărei cunoaştere este adesea deosebit de utilă. Se citează un exemplu, întîlnit în multe aplicaţii statistice: dacă o fabrică de confecţii studiază, printre altele, taliile cumpărătorilor, ea nu caută să fabrice „pantaloni medii", care nu se vor potrivi nimănui, ci „pantaloni pentru clasa modală", adică pentru „marea majoritate"; celelalte talii le va repartiza în raport de dispersia în jurul valorilor centrale [109, p. 221—222]. Ca şi în cazul medianei, intervalele marginale deschise ale unei repartiţii nu împiedică calculul modului.

Cu prilejul enunţării procedeului de estimare a modului prin formula (4.7), am observat că, dacă repartiţia este simetrică, media, mediana şi modul se confundă. Media aritmetică trebuie considerată suficientă pentru descrierea unei astfel de repartiţii. Cînd repartiţia se etalează către cele mai mari valori (curbă oblică la stînga), media aritmetică este valoarea centrală cea mai ridicată, modul cea mai mică, mediana se situează între cei doi parametrii. Tabelul 4.10 ne arată într-adevăr această ordine de mărime a parametrilor, care descriu repartiţia căsătoriilor pe grupe de vîrstă ale soţilor, de tip pronunţat asimetric (vezi exemplul 5).

Tabelul 4.10

		1970	1971	1972	1973	1974
Vîrsta soţului (ani)	medie ($\bar{x}$)	26,43	26,17	26,05	26,00	26,05
	mediană (Me)	23,99	23,77	23,74	23,69	23,76
	modală (*Mo*)	22,78	22,66	22,65	22,63	22,68
Vîrsta soţiei (ani)	medie ($\bar{x}$)	22,51	22,40	22,41	22,40	22,49
	mediană (*Me*)	20,56	20,67	20,79	20,85	20,94
	modală (*Mo*)	19,19	19,69	20,05	20,19	20,25

Notă. Calcule efectuate pe baza repartiţiilor de frecvenţă publicate în *Anuarul statistic al R.S. România 1972* şi *1975*.

În cazul repartiţiilor asimetrice dreapta, media aritmetică are mărimea cea mai scăzută, modul — cea mai ridicată, iar media se situează, de asemenea, între cei doi parametri. De exemplu, pentru repartiţia uşor asimetrică din tabelul 4.9, s-au obţinut:

$\bar{x} = 12{,}41$ lei; Me = 12,48 lei; Mo = 12,55 lei.

Exceptînd situaţiile care motivează clar alegerea unui singur parametru, utilizarea împreună a celor trei parametri oferă o mai completă reprezentare a realităţii. Repartiţiile asimetrice impun oricum calculul modului, în afară de cel al mediei aritmetice şi al medianei.

Trebuie să accentuăm încă o dată că obţinerea cu aproximaţie bună a parametrilor de nivel, pentru repartiţiile pe intervale, depinde mult de scara aleasă.

Scanned with OKEN Scanner

Un exemplu concludent: dimensiunile medii ale întreprinderii calculate pe baza repartiţiilor din tabelul 4.4 diferă simţitor de cele calculate din date directe — numărul total al muncitorilor şi al întreprinderilor pentru fiecare an:

Tabelul 4.11

Anii	Numărul de muncitori, în medie pe o întreprindere, calculat din:	
	date individuale	date grupate
1950	483	505
1974	1501	1451

Repartiţiile construite pe o scară cu intervale prea mari şi, în acelaşi timp, deschise la ambele capete explică diferenţele mari între mediile obţinute.

Mediana şi mai ales modul sînt şi ele influenţate în mare măsură de felul cum este alcătuită scara repartiţiei. Simpla deplasare cu cîteva unităţi poate modifica, de pildă, valoarea modală *.

Tabelul 4.12

x	f'	f''
14—16	4	25
16—18	21	
18—20	27	60
20—22	33	
22—24	39	80
24—26	41	
26—28	32	56
28—30	24	
30—32	10	17
32—34	7	
34—36	4	6
36—38	2	

* Vezi tabelul 4.3
$M'o = 24{,}49$, $M''o = 23{,}8$

4.2.2 Parametrii dispersiei

Media, mediana, modul ne informează, fiecare în felul său, despre ordinul de mărime mijlociu sau valoarea centrală a unei repartiţii statistice. Orice repartiţie are însă o organizare interioară definită de modul în care valorile variabilei gravitează în jurul valorii centrale şi care poate diferi de la o repartiţie la alta. Ca atare, parametrii de nivel trebuie neapărat completaţi cu parametrii de dispersie [18].

[18] Parametrii de nivel corespund punctelor celor mai remarcabile ale curbei de frecvenţe, iar parametrii de dispersie — etalării curbei. Alura intrinsecă a curbei de frecvenţe şi poziţia sa în raport cu axele de referinţă se reduc la un mic număr de mărimi, analog definirii unui cerc prin mărimile care fixează poziţia centrului şi raza.

Scanned with OKEN Scanner

Exemplul următor clarifică distincţia între cele două categorii de parametri; el priveşte repartiţiile ipotetice ale retribuţiilor orare în trei grupe de muncitori (vezi exemplul 6).

Tabelul 4.13

Retribuţia medie orară (lei)	Număr de muncitori		
	A	B	C
8—9	7		
9—10	12	8	
10—11	28	11	5
11—12	39	51	10
12—13	82	97	47
13—14	47	63	114
14—15	21	14	58
15—16	9	6	12
16—17	5		4
Total	250	250	250

Confruntarea repartiţiilor, două cîte două, arată:

A cu B — valorile centrale se situează în acelaşi interval „12—13 lei"; dar în timp ce retribuţiile individuale se eşalonează de la 8 lei la 17 lei în repartiţia A, ele se cuprind între 9 lei şi 16 lei în repartiţia B (dispersii diferite);

B cu C — dispersiile sînt aproximativ echivalente, dar gravitează în jurul unor valori centrale diferite;

A cu C — şi valorile centrale, şi dispersiile diferă.

Ca şi în cazul tendinţei sau valorii centrale, statisticienii au imaginat mai mulţi parametri pentru măsura dispersiei, întemeiaţi, în general, pe noţiunea de *abatere*. Parametrul cel mai simplu, dar şi cel mai rudimentar, este *amplitudinea repartiţiei* (ω), adică abaterea între cea mai mare (x_{max}) şi cea mai mică (x_{min}) valoare observată. Repartiţiile din tabelul 4.13 au amplitudinile: $\omega_A = 9$ lei; $\omega_B = 7$ lei şi $\omega_C = 7$ lei.

Suspectată pentru că este afectată de valori excepţionale, de fluctuaţii de selecţie mari de la eşantion la eşantion şi că nu ţine seamă de forma de repartiţie între limitele de variaţie, *amplitudinea* îşi găseşte utilizarea sa principală în controlul statistic al calităţii produselor (la serii de valori puţin numeroase).

Un parametru mai elaborat şi hotărît mai semnificativ în descrierea dispersiei se consideră cîteodată *abatere cuartilă*. Cuartilele sînt valorile care separă repartiţia în patru părţi, cu acelaşi număr de observaţii. Ca mărime, prima cuartilă (Q_1) este inferioară medianei, a doua cuartilă (Q_2) se confundă cu mediana, iar a treia cuartilă (Q_3) este superioară medianei [19]. Poartă numele

[19] Noţiunea de mediană se prelungeşte prin noţiunile de cuartile, decile, centile. *Decilele*, în număr de 9, împart cîmpul de variabilitate în 10 intervale, conţinînd fiecare 10% din observaţii. A cincea decilă se confundă, desigur, cu mediana. *Centilele*, în număr de 99, separă repartiţia în o sută de intervale egale; cele două centile extreme determină un interval care conţine 98% din observaţii. Decilele şi centilele sînt rareori utilizate pentru seriile care comportă destul de numeroase elemente, în prealabil aşezate în ordine crescătoare.

Scanned with OKEN Scanner

de *abatere cuartilă* (abatere intercuartilă sau simplu intercuartilă) distanța între prima și a treia cuartilă. Analog medianei, cuartila inferioară se calculează după formula:

$$Q_1 = x_k + d_k \frac{1/4(\Sigma f + 1) - F_{k-1}}{f_k} \tag{4.8}$$

și cuartila superioară

$$Q_3 = x_k + d_k \frac{3/4\,(\Sigma f + 1) - F_{k-1}}{f_k}. \tag{4.9}$$

Aplicate la repartiția din tabelul 4.9:

$$Q_1 = 11 + 1 \frac{1/4\,(250 + 1) - 47}{39} = 11{,}40 \text{ lei}$$

și

$$Q_3 = 13 + 1 \frac{3/4\,(250 + 1) - 168}{47} = 13{,}43 \text{ lei},$$

de unde abaterea intercuartilă (I)

$$I = Q_3 - Q_1 = 13{,}43 - 11{,}40 = 2{,}03 \text{ lei}.$$

Uneori, se utilizează *semiintercuartila* (sau amplitudinea semiintercuartilă):

$$\frac{Q_3 - Q_1}{2} = \frac{13{,}43 - 11{,}40}{2} = 1{,}02 \text{ lei}.$$

În sfîrșit, pentru anumite comparații (de exemplu, repartiții exprimate în unități diferite) se recurge *la abaterea intercuartilă relativă*, definită de relația:

$$\frac{Q_3 - Q_1}{Q_2} = \frac{13{,}43 - 11{,}40}{12{,}48} = 0{,}16,$$

adică de raportul dintre abaterea intercuartilă în valoare absolută și mărimea mediană. Evident, cu cît abaterea intercuartilă, în valoare absolută sau în valoare relativă, este mai mare, cu atît observațiile sînt mai dispersate și invers.

Cuartilele pot fi determinate cu o oarecare aproximație prin procedeul grafic, operîndu-se la fel ca în cazul valorii mediane (fig. 4.12).

Cu toate că abaterea intercuartilă prezintă avantajul posibilității de a fi aplicată deopotrivă la repartiții închise ca și la repartiții deschise și, în al doilea rînd, nu este prea mult influențată de valorile extreme, ea rămîne totuși un parametru insuficient pentru măsurarea dispersiei, întrucît nu-i expresia întregii repartiții. Un bun parametru de dispersie trebuie să integreze într-o valoare de sinteză abaterile tuturor valorilor observate între ele sau față de medie.

Diferența medie. Statisticianul italian Corrado Gini a propus drept măsură a dispersiei „media aritmetică a diferențelor în valoare absolută, formată prin asocierea observațiilor două cîte două în toate chipurile posibile, inclusiv cu ele însele", pe care a numit-o *diferența medie*.

Scanned with OKEN Scanner

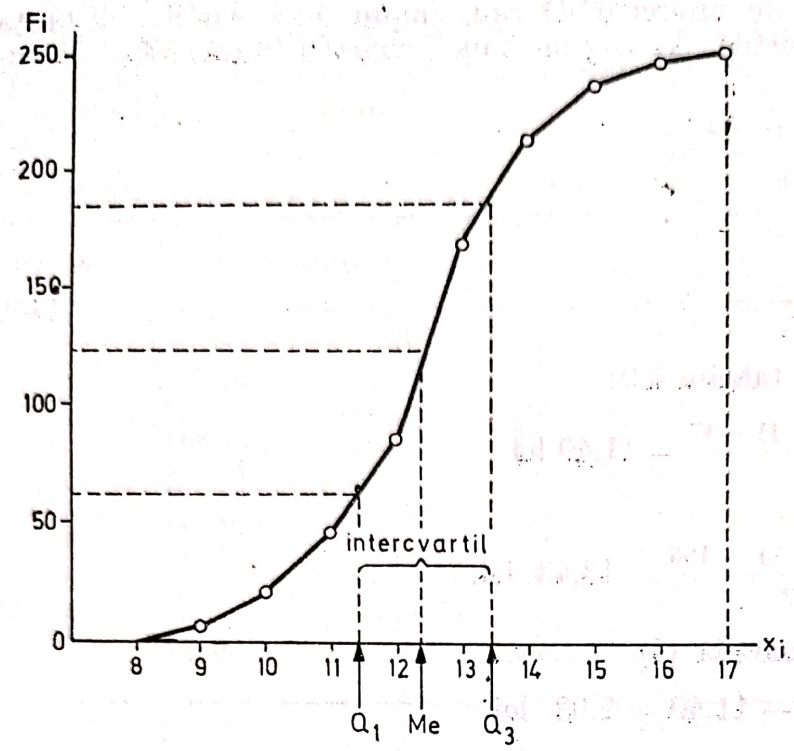


Fig. 4.12 Determinarea grafică a cuartilelor.

Pentru exemplificarea tehnicii de calcul, să folosim mai întîi seria scurtă de la p. 139 (nota 16) şi să convenim a înlocui „intrusul" cu valoarea mediană. Comodă este construirea unui tabel triunghiular: pe rîndul 1 se scriu valorile seriei, iar pe rîndurile următoare, diferenţele între fiecare valoare şi celelalte valori, începînd cu prima valoare [20].

Tabelul 4.14

Valorile seriei	103	115	119	128	137	138	143	Suma diferenţelor
	0	12	16	25	34	35	40	162
		0	4	13	22	23	28	90
			0	9	18	19	24	70
				0	9	10	15	34
					0	1	6	7
						0	5	5
							0	0
								368

Dacă se socotesc diferenţele într-un singur sens, ca în tabelul adoptat, atunci diferenţa medie are valoarea:

$$D = \frac{368}{21} = 17{,}52.$$

[20] A se vedea A. Piatier [109, p. 231—233].

Scanned with OKEN Scanner

Se observă că numărul total al diferenţelor este egal cu $1 + 2 + 3 + \ldots + (n - 1)$ pentru o serie de n elemente, adică expresia sumei unei progresii aritmetice de n termeni, cu primul termen 1 şi cu raţia 1. Suma sa va fi:

$$\Sigma(n - 1) = \frac{n(n-1)}{2} = \frac{7(7-1)}{2} = 21.$$

Aplicarea integrală a definiţiei diferenţei medii, reprodusă mai sus, înseamnă să se socotească diferenţele termenilor seriei în ambele sensuri (de exemplul 103—115, dar şi 115—103) şi de asemenea diferenţele nule (103—103; 115—115; etc.) Astfel [21]:

$$D = \frac{2 \times 368}{(7)^2} = \frac{736}{49} = 15{,}02 \cdot$$

C. Gini a conceput o formulă echivalentă:

$$D = 4 \frac{\Sigma | x - Me | \, | y_x - y_{Me} |}{n^2}, \qquad (4.10)$$

în care x reprezintă termenii seriei;

Me — mediana;
y_x — rangul fiecărui termen al seriei;
y_{Me} — rangul median;
n — numărul termenilor seriei.

Cu datele seriei luată ca exemplu, se formează tabelul de calcul:

Tabelul 4.15

x	y_x	$\lvert x - Me \rvert$	$\lvert y_x - y_{Me} \rvert$	$\lvert x - Me \rvert \, \lvert y_x - y_{Me} \rvert$
103	1	25	3	75
115	2	13	2	26
119	3	9	1	9
128	4	0	0	0
137	5	9	1	9
138	6	10	2	20
143	7	15	3	45
Total	—	—	—	184

Diferenţa medie :

$$D = 4 \frac{184}{(7)^2} = \frac{736}{49} = 15{,}02.$$

[21] În loc de un tabel triunghiular se obţine un tablou pătrat cu 7 coloane şi 7 rînduri (49 diferenţe) şi suma diferenţelor evident dublă (2 × 368).

Scanned with OKEN Scanner

În cazul unei repartiţii cu intervale egale, formula lui Gini devine:

$$D = 2d\frac{\Sigma F_i(F_i - \Sigma f)}{(\Sigma f)^2}, \qquad (4.11)$$

în care

d este mărimea intervalului;
f — frecvenţe simple;
F — frecvenţe cumulate.

Să calculăm diferenţa medie a repartiţiei din tabelul 4.9 aplicînd formula (4.11).

Tabelul 4.16

x	f	F	$F - \Sigma f$	$F(F - \Sigma f)$
8—9	7	7	243	1 701
9—10	12	19	231	4 389
10—11	28	47	203	9 541
11—12	39	86	164	14 104
12—13	82	168	82	13 776
13—14	47	215	35	7 525
14—15	21	236	14	3 304
15—16	9	245	5	1 225
16—17	5	250	0	0
Total	250	—	—	55 565

$$D = 2 \times 1 \frac{55\,565}{(250)^2} = 1{,}78 \text{ lei.}$$

Diferenţa medie a lui Gini are proprietatea de a depinde de abaterile valorilor variabilei însăşi şi nu de „răspîndirea valorilor în jurul unui punct arbitrar", cum ar fi o valoare centrală. Într-un anumit sens, ea măsoară „răspîndirea intrinsecă a populaţiei independent de originea de calcul sau de nivelul repartiţiei" [134, p. 165]. Această proprietate o recomandă ca parametru de dispersie pentru repartiţii statistice care nu au valoare centrală semnificativă [22].

Tendinţa centrală a unei repartiţii se exprimă în mod curent prin media aritmetică şi, prin urmare, este logic să se definească dispersia pe baza abaterilor măsurate în raport cu media aritmetică. Ca o măsură de expedient poate fi utilizată media abaterilor în valoare absolută, denumită abatere medie aritmetică, abatere absolută medie sau, pe scurt, *abatere medie*. Expresia sa teoretică este:

$$e_M = \frac{\Sigma |x - \bar{x}| f}{\Sigma f}. \qquad (4.12)$$

[22] De exemplu, în cazul repartiţiilor hiperbolice.

Scanned with OKEN Scanner

În exemplul precedent:

Tabelul 4.17

x	f	$\lvert x - 12{,}408 \rvert$	$\lvert x - 12{,}408 \rvert f$
8,5	7	3,908	27,356
9,5	12	2,908	34,896
10,5	28	1,908	53,424
11,5	39	0,908	35,412
12,5	82	0,092	7,544
13,5	47	1,092	51,324
14,5	21	2,092	43,932
15,5	9	3,092	27,828
16,5	5	4,092	20,460
Total	250	—	302,176

$$\bar{x} = 12{,}408$$

$$e_M = \frac{302{,}176}{250} = 1{,}21 \text{ lei.}$$

Abaterea medie nu poate fi supusă calculului algebric. Parametrul de dispersie, îndeobște preferat, rămîne abaterea pătratică medie.

Abaterea pătratică medie [23]. Se calculează pornind de la pătratele abaterilor, valori într-un fel mai semnificative decît abaterile simple în valoare absolută, căci accentuează efectul abaterilor mari, responsabile de dispersii mari. Formula ei de definiție este:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{\Sigma f}}. \tag{4.13}$$

Prin dezvoltarea formulei de definiție se ajunge la o formulă convenabilă în calculul practic:

$$s = \sqrt{\frac{\Sigma x^2 f - \frac{(\Sigma xf)^2}{\Sigma f}}{\Sigma f}}. \tag{4.14}$$

Formula (4.14) evită operația $(x - \bar{x})$, care poate provoca erori cînd media $(\bar{x})$ este complicată și comportă îndeosebi mai multe zecimale.

Calculul abaterii pătratice medii, pentru o repartiție cu intervale egale, se ușurează mult dacă se aplică formula:

$$s = \sqrt{\frac{\Sigma\left(\frac{x - x_0}{d}\right)^2 f}{\Sigma f} - \left[\frac{\Sigma\left(\frac{x - x_0}{d}\right) f}{\Sigma f}\right]^2} \cdot d. \tag{4.15}$$

[23] Numită și abatere standard (engl. *standard deviation*) sau abatere tip (fr. *écart-type*).

Scanned with OKEN Scanner

Se efectuează o schimbare de variabilă, procedeu al cărui principiu a fost expus în legătură cu calculul mediei aritmetice.

Mecanismul aplicării formulelor (4.14) și (4.15) este arătat în tabelul în care s-a reluat repartiția căsătoriilor după vîrsta soțului (tabelul 4.18).

Tabelul 4.18

Vîrsta x	Număr de căsătorii f	xf	x^2f	$\frac{x-37}{5}$	$\left(\frac{x-37}{5}\right)f$	$\left(\frac{x-37}{5}\right)^2 f$
0	1	2	3	4	5	6
17	8 592	146 064	2 483 088	−4	− 34 368	137 472
22	92 856	2 042 832	44 942 304	−3	−238 568	835 704
27	44 816	1 210 032	32 670 864	−2	− 89 632	179 264
32	12 516	400 512	12 816 384	−1	− 12 516	12 516
37	5 933	219 521	8 122 277	0	0	0
42	3 159	132 678	5 572 476	1	3 159	3 159
47	2 217	104 199	4 897 333	2	4 434	8 868
52	1 496	77 792	4 045 184	3	4 488	13 464
57	906	51 642	2 943 594	4	3 624	14 496
62	3 005	186 310	11 551 220	5	15 025	75 125
Total	175 496	4 571 582	130 044 744	—	−384 354	1 280 068

Potrivit datelor din col. 1—3, abaterea pătratică medie este egală cu:

$$s = \sqrt{\frac{130\,044\,744 - \frac{(4\,571\,582)^2}{175\,496}}{175\,496}} = 7{,}90 \text{ ani (formula 4.14)},$$

valoare confirmată și de calculul pe baza datelor din col. 4—6:

$$s = \sqrt{\left[\frac{1\,280\,068}{175\,496} - \left(-\frac{384\,354}{175\,496}\right)^2\right]} \cdot 5 = 7{,}90 \text{ ani (formula 4.15)}.$$

Pătratul abaterii pătratice medii se numește *varianță* (s^2). Tot așa cum media aritmetică ($\bar{x}$) este suma valorilor observate ponderate cu frecvențele lor [24], varianța (s^2) este suma, de asemenea ponderată cu frecvențele, a pătratelor abaterilor între valorile observate și media lor aritmetică. Definită astfel, ea se exprimă în unități bizare: acelea ale valorilor ridicate la pătrat. Ca să se rămînă la unități omogene, se reține, drept valoare tipică a dispersiei, rădăcina pătrată a varianței (s).

Noțiunile de varianță și abatere pătratică medie sînt fundamentale: se regăsesc explicit sau implicit formulate în aproape toate dezvoltările metodei statistice.

Coeficientul de variație. Parametrii de dispersie examinați se exprimă în unitatea de măsură a variabilei. Nivelurile medii ale variabilelor în raport cu care se calculează pot fi net diferite. De aceea, cînd trebuie să se compare dispersia mai multor repartiții, se recurge la un parametru de dispersie relativ.

[24] Frecvențe relative.

Scanned with OKEN Scanner

În aplicaţii, curent, se foloseşte *coeficientul de variaţie* (V), raportul, de regulă procentual, al abaterii pătratice medii faţă de media aritmetică:

$$V = \frac{s}{\bar{x}} \cdot 100. \tag{4.16}$$

Citǎm, cu titlu de exemplu, evoluţia celor trei parametri: media aritmetică ($\bar{x}$), abaterea pătratică medie (s) şi coeficientul de variaţie (V), în cazul repartiţiei căsătoriilor după vîrsta soţilor *.

Tabelul 4.19

Parametrii \ Anii		1970	1971	1972	1973	1974
$\bar{x}$ (ani)	soţ	26,43	26,17	26,05	26,00	26,05
	soţie	22,51	22,40	22,41	22,40	22,49
s (ani)	soţ	8,35	8,12	7,94	7,92	7,90
	soţie	8,03	7,78	7,62	7,54	7,54
V (%)	soţ	31,6	31,0	30,5	30,5	30,3
	soţie	35,7	34,7	34,0	33,7	33,5

* Vezi *nota* la tabelul 4.10.

Coeficientul de variaţie are aceeaşi semnificaţie ca şi abaterea pătratică medie: cu cît valoarea lui este mai mare cu atît mai eterogenă este populaţia pe care o caracterizează media, cu atît mai intensă este variaţia caracteristicii cercetate şi invers.

4.2.3 Caracteristicile formei

Despre forma unei repartiţii statistice ne putem da seama din examinarea curbei reprezentative. Dacă valorile variabilei sînt egal dispersate de o parte şi de alta a valorii centrale, avem de-a face cu o repartiţie simetrică, în caz contrar — cu o repartiţie asimetrică (disimetrică). O repartiţie este mai mult sau mai puţin aplatizată (sau boltită) după cum valorile corespunzătoare unei abateri mici de la valoarea centrală deţin o proporţie mai însemnată sau mai puţin însemnată.

S-au imaginat şi coeficienţi susceptibili să definească într-un mod foarte general forma repartiţiei. Cel mai des utilizaţi sînt coeficienţii γ_1, ai lui K. Pearson, în ce priveşte asimetria.

$$\gamma_1 = \frac{\sum (x - \bar{x})^3 f}{s^3 \cdot \sum f} \tag{4.1}$$

şi γ_2 a lui R. A. Fisher pentru caracterizarea boltirii

$$\gamma_2 = \frac{\sum (x - \bar{x})^4 f}{s^4 \cdot \sum f} - 3.$$

Scanned with OKEN Scanner

Cei doi coeficienţi se pot calcula uşor pornind de la aceleaşi transformări de variabilă efectuate în calculul mediei aritmetice şi abaterii pătratice medii. Să ne referim, din nou, la repartiţia întreprinderilor agricole de stat după producţia medie la hectar de grîu—secară (tabelele 4.3 şi 4.20).

Tabelul 4. 20

x	f	x'	$x'f$	$(x')^2f$	$(x'-\bar{x}')$	$(x'-\bar{x}')^3$	$(x'-\bar{x}')^3f$	$(x'-\bar{x}')^4$	$(x'-\bar{x}')^4f$
1500	4	−6	−24	144	−4,5082	−91,62	−366,48	413,04	1652,16
1700	21	−5	−105	525	−3,5082	−43,18	−906,78	151,48	3181,08
1900	27	−4	−108	432	−2,5082	−15,78	−426,06	39,58	1068,66
2100	33	−3	−99	297	−1,5082	−3,43	−113,19	5,17	170,61
2300	39	−2	−78	156	−0,5082	−0,13	−5,07	0,07	2,73
2500	41	−1	−41	41	+0,4918	+0,12	+4,92	0,06	2,46
2700	32	0	0	0	0	0	0	0	0
2900	24	+1	24	24	+2,4918	+15,47	+371,28	38,55	925,20
3100	10	+2	20	40	+3,4918	+42,57	+425,70	148,65	1486,50
3300	7	+3	21	63	+4,4918	+90,63	+634,41	407,09	2849,63
3500	4	+4	16	64	+5,4918	+165,63	+662,52	909,61	3638,44
3700	2	+5	10	50	+6,4918	+273,59	+547,18	1776,09	3552,18
Total	244	—	−364	1836	—	—	+828,43	—	18529,65

$$s = \sqrt{\frac{1\,836 - \frac{(-364)^2}{244}}{244}} = 2{,}31$$

$$\gamma_1 = \frac{828{,}43}{244 \cdot (2{,}31)^3} = 0{,}2754$$

$$\gamma_2 = \frac{18\,529{,}65}{244 \cdot (2{,}31)^4} - 3 = 2{,}67 - 3 = 0{,}33.$$

Repartiţia are o uşoară asimetrie, iar boltirea suferă o oarecare distorsiune de la curba normală.

4.2.4 Curba de concentrare

În cercetările statistice, noţiunea de concentrare este înrudită cu cea de dispersie, întrucît ea se raportează la intensitatea grupării datelor în jurul valorii centrale. Gradul de concentrare poate fi apreciat şi măsurat cu ajutorul curbei de concentrare, o curbă sintetică ţinînd seama simultan de numărul observaţiilor şi suma valorilor corespunzătoare fiecărui interval al repartiţiei.

Curba de concentrare se construeşte pe baza frecvenţelor relative cumulate. Tabelul 4.21 cuprinde datele pentru construirea curbei de concentrare a industriei de subordonare republicană.

Scanned with OKEN Scanner

Tabelul 4.21

Procente

Grupe de întreprinderi după numărul de muncitori	Numărul întreprinderilor		Numărul muncitorilor	
	Frecvenţe simple	Frecvenţe cumulare	Frecvenţe simple	Frecvenţe cumulate
Pînă la 200	13,1	13,1	1,0	1,0
201—500	15,2	28,3	3,5	4,5
501—1000	22,2	50,5	11,1	15,6
1001—2001	26,0	76,5	24,7	40,3
2001—3000	12,2	88,7	19,4	59,7
3001—5000	7,7	96,4	18,8	78,5
Peste 5000	3,6	100,0	21,5	100,0

Sursa: *Anuarul statistic al Republicii Socialiste România, 1975*, p. 86—87.

Se pun în corespondenţă două serii de date: procentul cumulat al numărului de întreprinderi şi procentul cumulat al numărului de muncitori (fig. 4.13). Suprafaţa mărginită de curba de concentrare şi diagonala AC poartă numele de *arie de concentrare:* cu cît această arie este mai importantă, cu atît concentrarea este mai puternică. La limită, (în ipoteza teoretică de concentrare maximă) curba de concentrare s-ar confunda cu linia frîntă ABC. Şi dacă toate observaţiile ar fi egale (20% de întreprinderi ar corespunde la 20% de muncitori; 30% la 30% etc.), curba de concentrare s-ar reduce la diagonala AC, numită uneori dreaptă de echilibru sau de echirepartiţie [25].

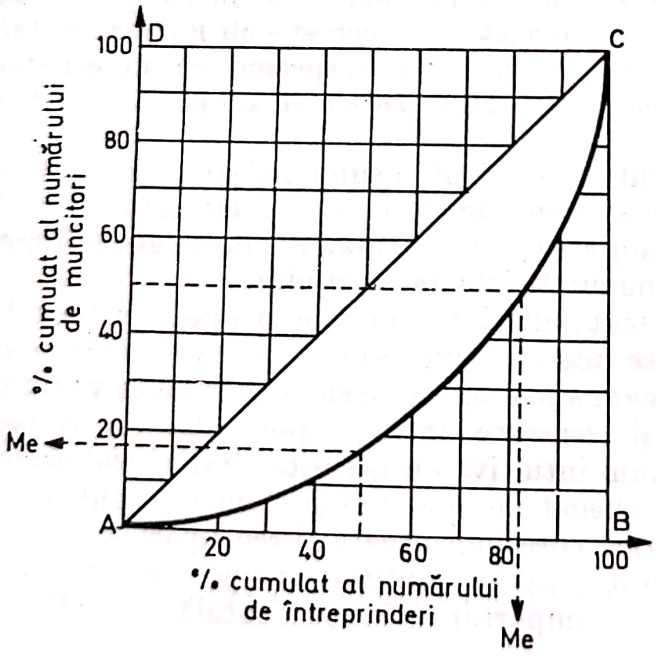


Fig. 4.13 Curba de concentrare a întreprinderilor.

[25] Reprezentarea grafică a concentrării a fost pusă la punct de M. O. Lorenz şi C. Gini. S-a propus şi un indice de concentrare (C):

$$C = \frac{\text{Aria de concentrare}}{\text{Aria triunghiului } ABC}.$$

CAPITOLUL 5

Sondajul statistic (I). Metode de estimație și de verificare a ipotezelor statistice. Eșantioane mari

5.1 EȘANTIOANE. POPULAȚIE ORIGINARĂ

Teoria modernă a sondajului statistic s-a constituit și dezvoltat sub impulsul nevoilor practice. Într-adevăr, la rezolvarea a numeroase probleme practice, observatorii au fost confruntați cu necesitatea de a studia populații numeroase fără a avea însă posibilitatea de a examina fiecare element component al acestora. De pildă, este evident că o întreprindere industrială modernă, automatizată, producînd în masă diferite tipuri de produse industriale nu are posibilitatea reală de a constata calitatea producției pe care o realizează, examinînd în acest scop fiecare produs în parte, în timpul procesului de fabricație sau în stadiul de recepție, la beneficiar; același lucru se întîmplă și la controlul calității recoltei agricole sau la determinarea volumului ei probabil etc. În asemenea probleme s-a considerat că cel mai bun lucru pe care-l poate face cel interesat să le rezolve este acela de a cerceta un număr limitat de elemente — o *parte* din ansamblul populației —, sperînd că informația primită pe această cale îi va fi utilă și, totodată, suficient de precisă pentru a cunoaște întreaga populație.

Acest număr limitat de elemente — de fapt, o subpopulație — constituie ceea ce în teoria sondajului se cunoaște sub denumirea de *eșantion* — concept ce se îmbină dialectic cu cel de *populație originară* — indicînd întreaga mulțime, finită sau infinită, a elementelor populației luate în studiu.

Un postulat unanim admis, bazat inițial pe intuiție și apoi verificat în numeroase cercetări aplicative este acela că un eșantion ne va spune ceva util despre populația originară din care a fost extras și că pe măsură ce volumul eșantionului crește, *proprietățile* și *structura* întregii populații vor fi mai fidel relevate. Toate aceste accepțiuni intuitive au necesitat însă și elaborarea unui suport logic, a unui sistem închegat de aprecieri și estimații cantitative ca și a unor generalizări matematice corespunzătoare. Preocupări și direcții de cercetare tot mai ample în acest sens au conturat și apoi au cristalizat *obiectul teoriei sondajului*, capitol de importanță fundamentală al statisticii matematice moderne.

Tipuri de populații originare. În teoria sondajului s-au constituit cîteva tipuri distincte de populații, generalizate în concepte de circulație generală în literatura de specialitate cum sînt: populații *finite*, populații *infinite* și populații *ipotetice*.

„Populația" locuitorilor din R. S. România, „populația" întreprinderilor industriale de stat sau „populația" cărților de la Biblioteca de stat etc. sînt

Scanned with OKEN Scanner

exemple de populaţii finite; ele sînt, totodată, şi populaţii *reale*, în sensul că sînt constituite dintr-un număr — efectiv şi concret — de elemente. „Populaţia" mărimilor posibile ale producţiei globale industriale, ale recoltei, ale presiunilor atmosferice şi, în general, valorile pe care le poate lua o caracteristică cantitativă *continuă*, sînt tot atîtea exemple de populaţii infinite. În multe cazuri, numărul elementelor unei populaţii este atît de mare încît practic o asemenea populaţie o putem considera infinită. Studiul populaţiilor infinite este adeseori mai simplu decît cel al populaţiilor finite şi de aceea numeroase probleme practice pot fi abordate şi cercetate, făcînd ipoteza infinităţii populaţiei, fără ca prin aceasta să se facă o eroare sensibilă.

Printr-o extindere logică a conceptului de populaţie reală s-a constituit şi conceptul de *populaţie ipotetică.*

Dacă ne referim la cazul clasic al experimentelor de aruncare a zarului, este evident că putem să ne imaginăm o infinitate de asemenea experimente şi, astfel, toate aruncările posibile ale zarului ne sugerează ideea unei populaţii ipotetice. Formal, o asemenea populaţie o putem defini ca mulţime a tuturor modurilor posibile ce pot fi concepute şi prin care se poate produce un eveniment specificat. De fapt, o populaţie ipotetică poate fi imaginată în legătură cu orice eveniment observat. De pildă, ne putem imagina populaţia valorilor posibile pe care le va putea lua producţia agricolă în următorii cinci ani sau populaţia parităţilor valutare de pe piaţa internaţională în perspectiva anului 2 000 etc. Nu avem decît să ne imaginăm toate circumstanţele, înainte ca evenimentul să se producă şi, astfel, populaţia noastră ipotetică va fi formată din toate căile posibile pe care evenimentul s-ar putea produce.

Introducerea în ştiinţă a conceptului de populaţie ipotetică reprezintă un important pas înainte în cercetarea modelelor teoretice ale diferitelor fenomene din natură şi societate, inclusiv ale celor economice.

Problemele care apar în contextul cercetării unor asemenea populaţii prezintă un pronunţat interes practic. Ca şi în cazul celorlalte populaţii, teoria şi tehnica sondajului ne poate oferi instrumente eficiente de lucru şi în acest domeniu.

Am arătat mai sus că metoda sondajului poate fi aplicată cu succes şi în studiul populaţiilor care, practic, pot fi considerate ca avînd un număr infinit de unităţi. Astfel, populaţia formată din globulele de diferite categorii ale sîngelui uman, populaţiile formate din particulele elementare ale corpurilor din fizica statistică sau populaţiile care cuprind rezervele subterane de minereuri, ţiţei etc., rezervele de material lemnos de diferite esenţe ale fondului silvic al ţării, nu se pot stabili printr-o observare totală, ci prin aplicarea sondajelor ştiinţific organizate.

5.2 AVANTAJELE ŞI OBIECTIVELE SONDAJULUI

Metoda sondajului se bucură de privilegiul de a fi adeseori singura metodă de cercetare practic posibilă. Ne referim, desigur, la împrejurările în care înregistrarea *totală* nu se poate utiliza fără prejudiciul ireparabil adus obiectelor studiate. De pildă, acest lucru se întîmplă la controlul statistic al calităţii producţiei cînd se pun probleme privind măsurarea duratei medii de ardere a becurilor, încercarea rezistenţei la rupere a firelor etc.

Scanned with OKEN Scanner

Obiectivul fundamental al cercetării, pe bază de sondaje, a populațiilor statistice este acela *de a obține cu efort minim un volum maxim de informație asupra populației originare luate în studiu.*

Avînd în vedere acest obiectiv, cercetătorul trebuie să urmeze un plan riguros de organizare a cercetării, alegînd metodele cele mai adecvate, în funcție de tipul informației pe care vrea s-o aibă la dispoziție. Cercetînd pe bază de sondaj o anumită populație el are în vedere, de obicei, una sau mai multe caracteristici cantitative sau anumite caracteristici-atribut (calitative), mărginindu-se, de pildă, la obiectivul de a ști *repartiția de frecvență* a caracteristicilor respective, pentru întreaga populație luată în studiu. Desigur, ar fi ideal dacă cel ce cercetează ar putea reprezenta această repartiție prin una din legile teoretice de repartiție expuse în capitolul 3. În multe cazuri însă este posibil ca populația originară să nu admită o astfel de reprezentare sau ca eșantionul să nu fie suficient de mare pentru ca cercetătorul să-și asume riscul de a-i acorda încredere. În asemenea cazuri el încearcă să găsească anumite *estimații* pentru diferite valori tipice ale întregii populații [1]. În funcție de scopul cercetării, determinarea acestor estimații poate fi pe deplin suficientă.

În general, se poate afirma că o bună parte a teoriei sondajului este consacrată găsirii procedeelor adecvate de estimație a valorilor tipice ale populației originare, pe baza informației oferite de eșantioane. Este de reținut un aspect esențial al teoriei estimației: deducțiile pe bază de eșantion nu sînt categoric deterministe, ci se formulează cu un anumit grad de *încredere* probabilistică. De aceea, un alt obiectiv al cercetării pe bază de sondaj este acela de a determina *precizia* unei estimații, stabilindu-se în acest scop anumite *limite* pentru abaterea posibilă a estimației bazate pe eșantion, față de valoarea adevărată a parametrului estimat al populației originare.

5.3 TIPURI DE SONDAJ

Un eșantion se formează dintr-un număr determinat de unități elementare de observare. După modul în care pot fi prelevate în eșantion aceste unități, s-au constituit cîteva tipuri diferite de sondaj. În acest paragraf vom expune doar unele probleme generale ale acestor categorii de sondaj, urmînd ca specificul și formele lor de organizare, metodele respective ale estimației și verificării ipotezelor statistice să fie dezvoltată în paragrafele următoare, pe marginea unor exemplificări concrete.

5.3.1 Sondaj pur aleator

În cazul acestui tip de sondaj, unitățile elementare ale populației originare sînt prelevate în eșantion, „la întîmplare". Formal, alegerea „la întîmplare" înseamnă că *fiecare element component al populației originare are aceeași probabilitate de a fi ales în eșantion.*

[1] Valorile tipice ca mediile, abaterile medii pătratice, momentele, mediana, modul, coeficienții de corelație ș.a.m.d. se mai numesc și *parametrii* populațiilor originare, spre deosebire de variabile corespunzătoare de sondaj care mai poartă denumirea de *statistici*.

Scanned with OKEN Scanner

Am formulat astfel un principiu fundamental al sondajului aleator. În mod analog vom vorbi şi de un *eşantion aleator*, cînd acesta a fost ales astfel încît, în procesul sondajului, toate eşantioanele posibile, formate, de pildă, din cîte n elemente aveau aceeaşi probabilitate de a se realiza.

Reprezentativitatea sondajului. O primă problemă care apare în procesul sondajului aleator este aceea a *modului* în care putem obţine un eşantion cu adevărat aleator, pentru a asigura astfel *reprezentativitatea sondajului.* Prin această noţiune înţelegem o anumită identitate — mai mare sau mai mică — a *structurii* eşantionului cu structura populaţiei originare. Obţinem o reprezentativitate *perfectă* a unui eşantion, cînd identitatea dintre cele două structuri este deplină. De obicei, se consideră satisfăcător acel sondaj care asigură abateri ale valorilor tipice de structură ale populaţiei de sondaj cu cel mult 5% faţă de parametrii corespunzători ai populaţiei originare.

Erori de sondaj. Prin erori de sondaj se înţeleg erori de reprezentativitate care pot fi *sistematice* sau *întîmplătoare.*

Prin erori de reprezentativitate se înţeleg diferenţele dintre caracteristicile studiate în eşantion şi cele din populaţia originară. Aceste diferenţe se produc în toate cazurile, eliminarea lor completă nefiind posibilă. Este posibilă însă determinarea cu exacticitate a mărimii lor probabile.

Vom sublinia că erorile de reprezentativitate — li se mai spune *deplasări* ale eşantionului — sînt, în general, rezultatul încălcării principiului fundamental al sondajului aleator pe care l-am formulat mai sus — acela ca fiecărei unităţi din populaţia generală să-i fie asigurată aceeaşi probabilitate de a fi aleasă în eşantion. Asemenea deplasări pot avea loc din multe cauze, printre care, de pildă, instrumentele de măsură imperfecte, tehnica defectuoasă de alegere, dar şi calităţile personale ale observatorului sau intervenţiile premeditate sau nepremeditate ale acestuia. Dacă ne referim la această din urmă cauză vom pune în evidenţă faptul — observat adeseori în practică — că fiinţa umană are, ca parte componentă a structurii sale psihologice, tendinţa de a se îndepărta de caracterul cu adevărat întîmplător în alegerile sale.

În teoria sondajului s-au elaborat o serie de procedee şi tehnici de sondaj care asigură — pe cît este cu putinţă în fiecare caz dat — caracterul aleator al sondajului şi ele sînt diferenţiate în funcţie de volumul eşantionului şi natura populaţiei. Este de relevat că anumite procedee care sînt potrivite în cercetarea populaţiilor mici se dovedesc cu totul ineficiente în cazul populaţiilor mari.

În general, se consideră că putem asigura, în fiecare caz dat de sondaj, un eşantion aleator dacă procedeul de alegere acceptat este independent de caracteristicile populaţiei pe care o cercetăm. Astfel, putem spune, în măsura în care ne referim la aceste caracteristici, că nu există nici o raţiune ca un element să fie ales mai curînd decît altul şi în consecinţă se va respecta principiul conform căruia fiecare element al populaţiei va avea aceeaşi şansă de a fi ales în eşantion. În acest sens, au fost concepute procedee cum ar fi, de pildă, cel al alegerii mecanice şi, mai recent, cel al tabelelor de numere întîmplătoare.

Scanned with OKEN Scanner

5.3.2 Sondaj dirijat

Un asemenea sondaj se efectuează după un anumit principiu prestabilit, atunci cînd acesta apare ca fiind raţional şi util. Lăsînd loc opiniilor personale şi inclinaţiilor proprii la extragerea eşantionului, ceea ce — după cum ştim — prezintă pericolul unor deplasări mari, încrederea pe care o putem acorda acestui tip de sondaj depinde de circumstanţele cazului dat.

Cu o anumită plauzibilitate se poate spune că uneori sondajul dirijat furnizează eşantioane tipice sau reprezentative într-o măsură mai mare decît un sondaj aleator. Nu este exclus ca un eşantion aleator să se abată mult faţă de medie, în timp ce un eşantion dirijat, prin construcţia sa, să fie foarte apropiat de medie.

Este de observat totuşi că, în cazul eşantioanelor mari, datorită deplasărilor, rezultatele sondajului dirijat pot fi din ce în ce mai puţin reprezentative, în timp ce rezultatele sondajului aleator să capete o reprezentativitate tot mai înaltă.

În practică, avantajele sondajului pur aleator şi ale celui dirijat pot fi util combinate, organizîndu-se un sondaj mixt.

5.3.3 Sondaj mixt

Cel mai reprezentativ exemplu de sondaj mixt este sondajul stratificat (tipic). El combină principiile sondajului dirijat cu cele ale sondajului aleator. Esenţa acestui tip de sondaj constă în divizarea întregii populaţii în straturi (grupe tipice în raport cu un principiu oarecare, prestabilit) şi apoi extragerea a cîte unui eşantion aleator din fiecare strat.

Să presupunem, de pildă, că ne interesează producţia agricolă globală lunară în întreprinderile agricole de stat din R. S. România, în care scop se consideră raţională şi utilă folosirea unui sondaj mixt. Este clar că, în faţa unei astfel de sarcini, toate I.A.S. vor fi clasificate în grupe tipice cum ar fi I.A.S.: „axate în principal pe culturi agricole", „axate în principal pe producţia de lapte", „axate în principal pe producţia de legume" ş.a.m.d. şi apoi din fiecare grupă, astfel formată, se va extrage cîte un eşantion aleator.

Şi în cazul sondajelor mixte apar deplasări; de regulă, ele sînt însă mai mici.

Sondajul mixt ne sugerează ideea că formarea unui eşantion reprezentativ se poate face în *faze* sau *secvenţial*. În terminologia teoriei sondajului figurează şi astfel de concepte. Explicaţii de detaliu asupra acestora vor fi date în paragrafele respective.

5.4 SONDAJ SIMPLU DIN POPULAŢII CU O SINGURĂ CARACTERISTICĂ CANTITATIVĂ

5.4.1 Generalităţi. Repartiţii de sondaj

Sondajul simplu este un caz particular al sondajului pur aleator.

În teoria şi practica acestui sondaj sînt admise următoarele ipoteze:

— fiecare unitate elementară are aceeaşi probabilitate p de a fi extrasă şi deci inclusă în eşantion;

Scanned with OKEN Scanner

— la fiecare extragere, fiecare din unităţile elementare poate fi considerată ca fiind extrasă din una şi aceeaşi sursă (aceeaşi populaţie originară);

— extragerea fiecărei unităţi elementare este independentă de oricare alta.

Experimentele clasice de aruncare a zarului, a monedei sau a extragerilor dintr-o urnă Bernoulli, cu revenirea bilei în urnă după fiecare extragere, pot servi drept modele de sondaj aleator simplu din populaţii ipotetice, practic infinite. Cazul urnei cu bila nerevenită oferă un model de sondaj, care nu mai este simplu, deşi este pur aleator.

Cele trei ipoteze formulate mai sus le admitem şi în cazul unui sondaj simplu, vizînd o caracteristică cantitativă oarecare, discretă sau continuă. În general, valorile variabilei, date de diferite experimente, vor fi cuprinse într-un interval, care poate fi nelimitat sau limitat pentru considerente de ordin practic sau teoretic.

Înainte de a prezenta trăsăturile esenţiale ale unui sondaj simplu de o singură variabilă vom introduce următoarele notaţii:

X sau X_i — simbol principal pentru o caracteristică cantitativă (variabilă) pe care o posedă unităţile din populaţia originară (în dezvoltările ulterioare cînd vom studia populaţii multidimensionale vor fi folosite şi alte litere majuscule: V, Z sau A, B);

x sau x_i — idem, pentru caracteristicile unităţilor din eşantion;

n — simbol pentru volumul eşantionului;

N — simbol pentru numărul total al elementelor populaţiei originare;

f — simbol pentru frecvenţa absolută;

F — frecvenţa absolută cumulată;

f/n — frecvenţa relativă în eşantion;

m sau $M(X)$ — media în populaţia originară;

σ — abaterea medie pătratică în populaţia originară;

s^2 — dispersia de sondaj;

$\tilde{s}^2$ — estimaţia dispersiei populaţiei originare;

s — abaterea medie pătratică de sondaj;

$\tilde{s}$ — estimaţia abaterii medii pătratice a populaţiei originare;

$\bar{x}$ — valoarea mediei aritmetice de sondaj;

$s_{\bar{x}}$ — eroarea standard a mediei de sondaj; acelaşi simbol se foloseşte şi pentru erori standard ale altor valori tipice, caz în care indicele de jos se referă la valoarea tipică respectivă de sondaj.

Notaţiile specifice pentru alte valori tipice de sondaj (mediană, momente etc.) vor fi introduse odată cu expunerea problemelor respective.

Pentru ilustrări practice poate fi considerată o reprezentare miniaturală a populaţiei originare, imaginîndu-ne că fiecărui element al acestei populaţii i-am ataşat un număr, scris pe o fişă, corespunzînd unei anumite valori a variabilei X. În acest caz, procesul de sondaj poate fi considerat ca o extragere de fişe, cu notarea valorilor lui X ataşate acestora. Variabila poate lua, în general, orice valoare într-o mulţime continuă şi mulţimea fişelor apare astfel ca fiind în mod real sau practic, infinită.

Scanned with OKEN Scanner

Cercetarea prin sondaj a unei caracteristici cantitative are un triplu obiectiv, şi anume:

— în primul rînd, de a *compara* observaţia de sondaj cu mărimea teoretică corespunzătoare populaţiei originare şi pe această bază a urmări în ce măsură abaterile ar putea fi atribuite fluctuaţiilor aleatoare ale sondajului;

— în al doilea rînd, de a *estima*, cu ajutorul eşantioanelor, unii parametri ai populaţiei originare;

— şi, în sfîrşit, de a *aprecia exactitatea* estimaţiilor efectuate.

Dacă luăm un număr mai mare de eşantioane dintr-o populaţie oarecare şi calculăm pentru fiecare eşantion în parte diferite valori tipice — media, abaterea standard, medii de ordin superior etc. — vom găsi, în general, serii de valori deosebite pentru fiecare din ele. Dacă numărul eşantioanelor este mare, aceste valori pot fi grupate într-o *repartiţie de frecvenţă* a valorii tipice de sondaj considerate; pe măsură ce numărul eşantioanelor creşte, această repartiţie va aproxima forma „ideală" a unei curbe. O astfel de repartiţie o vom numi *repartiţia de sondaj*.

5.4.2 Estimaţii în teoria eşantioanelor de mare volum

Pentru a expune sistematic estimaţiile posibile în teoria eşantioanelor mari, formate pe bază de sondaj simplu, vom începe printr-o cercetare aplicativă asupra unei repartiţii statistice reale.

Aplicaţia 1. Vom considera repartiţia întreprinderilor industriale de subordonare republicană, la sfîrşitul anului 1974, după numărul muncitorilor, ilustrată în tabelul 5.1.

Tabelul 5.1

REPARTIŢIA ÎNTREPRINDERILOR INDUSTRIALE DE SUBORDONARE REPUBLICANĂ DUPĂ NUMĂRUL MUNCITORILOR, ÎN 1974

Intervale de variaţie a numărului de muncitori* (X)	Mijlocul intervalului (X_i)	Frecvenţa absolută	
		de grupă (f_i)	cumulată (F_i)
200 — 500 muncitori	350	190	190
501 — 1000 muncitori	750	278	468
1001 — 2000 muncitori	1500	325	793
2001 — 3000 muncitori	2500	152	945
3001 — 5000 muncitori	4000	97	1042
Total	—	1042	—

Sursa: *Anuarul Statistic al R. S. România 1975*, p. 86

* Pentru a evita dificultăţile de calcul ce le implică o serie deschisă la ambele intervale marginale ca cea în cauză, am considerat în calcul numai întreprinderile de mărime mijlocie, în care amplitudinea de variaţie a numărului de muncitori este de la 200 la 5000. Această limitare nu afectează scopul didactic în care se va face calculul valorilor tipice de sondaj şi estimaţiile ce se vor face pe baza lor.

Scanned with OKEN Scanner

Vom calcula principalele valori tipice ale variabilei X pentru repartiția din tabelul 1, pentru ca, în continuare, folosind metoda sondajului simplu, să le estimăm pe baza unui eșantion aleator mare, format din 100 unități elementare, interpretînd și semnificația estimațiilor obținute.

Avem deci următoarele valori tipice principale în populația originară:

valoarea medie a numărului de muncitori:

$$M(X)=m=\frac{\sum_{i=1}^{5} x_i fi}{\sum_{i=1}^{5} fi}=\frac{(350\times 190)+(750\times 278)+(1500\times 325)+(2500\times 152)+(4000)}{190+178+325+152+97}=$$

$$=\frac{1530\,500}{1042}=1468{,}80998;$$

dispersia numărului de muncitori:

$$\sigma^2=\frac{\sum_{i=1}^{5}(X_i-m)^2 f_i}{\sum_{i=1}^{5} f_i}=\frac{11766528{,}528}{1\,042}=1119007{,}80.$$

Abaterea medie pătratică (abaterea standard) a numărului de muncitori:

$$\sigma=\sqrt{\sigma^2}=\sqrt{1119007{,}035}=1057{,}83128.$$

Estimarea mediei aritmetice, a dispersiei și abaterii medii pătratice, care caracterizează populația originară. Pentru a ilustra raționamentul statistic de estimare a parametrilor m, σ^2 și σ ai populației originare din 1042 de întreprinderi industriale republicane, de mărime mijlocie, avînd repartiția indicată în tabelul 5.1, vom presupune că parametrii respectivi nu ne sînt cunoscuți și că un eșantion mare[2] — format din 100 unități industriale, extras la întîmplare — trebuie să ne servească drept sursă de informație pentru estimarea lor.

Practic, acest eșantion l-am format cu ajutorul unui tabel de numere întîmplătoare[3]. Repartiția valorilor individuale cuprinse în eșantion este arătată în tabelul 5.2.

Tabelul 5.2

REPARTIȚIA DE SONDAJ A VALORILOR INDIVIDUALE CUPRINSE ÎN EȘANTION ($n = 100$)

x_i	f_i
350	20
750	28
1500	30
2500	13
4000	9
Total	100

[2] În cele ce urmează vom considera că eșantionul este mare dacă volumul lui este mai mare decît 30 ($n > 30$); justificarea teoretică o vom da cu prilejul analizei eșantioanelor mici, cu ajutorul repartiției t a lui Student.

[3] În anexa 10 este dat un tabel cu numere întîmplătoare care a fost folosit pentru a preleva eșantionul nostru aleator de 100 unități industriale.

Scanned with OKEN Scanner

Arătăm şi modul în care a fost folosit tabelul cu numere întîmplătoare la formarea eşantionului.

Din tabelul 5.1 ilustrînd repartiţia unui număr de 1042 întreprinderi industriale din ţara noastră la sfîrşitul anului 1974 — este vorba de „masivul de bază" al întreprinderilor noastre industriale de subordonare republicană (circa 80%) — ne-am pus problema de a extrage un eşantion aleator de 100 unităţi.

Populaţia originară a fost formată deci din 1042 elemente şi fiecărui element i-am ataşat un număr de ordine de la 1 la 1042. A alege un număr de 100 întreprinderi, la întîmplare, echivalează cu a găsi 100 de numere întîmplătoare între 1 şi 1042. Am luat, prin urmare, din fiecare coloană a tabelului cu numere întîmplătoare, succesiv, primele 100 de numere de cîte patru cifre mai mici decît 1042. Astfel am obţinut următoarele numere de ordine pentru elementele eşantionului nostru:

576,	877,	157,	199,	240,	93,	308,	654,	970,	298
773,	485,	158,	38,	402,	794,	1029,	537,	896,	230
591,	607,	169,	75,	265,	534,	631,	287,	677,	302
283,	981,	358,	129,	180,	862,	626,	462,	18,	56
714,	104,	904,	773,	428,	704,	559,	643,	85,	984
786,	678,	547,	365,	1026,	744,	541,	345,	950,	199
871,	815,	58,	44,	331,	375,	60,	419,	647,	1011
6,	784,	908,	133,	354,	29,	693,	703,	307,	293
869,	504,	423,	200,	997,	147,	788,	420,	866,	860
673,	870,	631,	172,	961,	350,	774,	852,	224,	410

Ultimelor cinci elemente ale eşantionului, li s-au ataşat după — cum se vede — numerele aleatoare 350, 774, 852, 224, şi 410 (vezi primele 4 cifre de pe ultima coloană a anexei nr. 10 la intersecţia cu rîndurile 67, 70, 74, 81 şi 95).

Dacă presupunem că numerotarea întreprinderilor industriale s-a făcut în ordinea crescătoare a numărului de muncitori, atunci se vede că prima întreprindere aleasă (cu numărul de ordine 576) aparţine grupei cu mijlocu intervalului de variaţie egal cu 1500 de muncitori (tabelul 5.1), a doua întrel prindere cuprinsă în eşantion (cu numărul de ordine 877) aparţine grupei cu mijlocul intervalului egal cu 2500 ş.a.m.d.

Pe baza celor 100 elemente extrase s-a constituit repartiţia de sondaj ilustrată în tabelul 5.2.

Estimarea mediei aritmetice. Estimarea acestei valori tipice în populaţia originară — valoare adevărată — o putem face cu ajutorul mediei aritmetice de sondaj, calculată pe baza datelor de sondaj din tabelul 5.2 şi întrucît eşantionul este mare, este foarte probabil ca valoarea de estimaţie să difere nesemnificativ de valoarea adevărată. În cazul exemplului nostru obţinem următoarea estimaţie a *mediei:*

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i} = \frac{141\,500}{100} = 1415.$$

Am găsit astfel media sondajului egală cu 1415, faţă de media populaţiei originare, care este egală cu 1468,80.

Scanned with OKEN Scanner

Această valoare trebuie considerată ca fiind o variabilă aleatoare în sensul că extrăgînd un alt eşantion aleator de 100 elemente sau chiar serii de mai multe eşantioane de cîte 100 unităţi elementare, vom obţine pentru fiecare eşantion o altă valoare medie de sondaj. Toate aceste valori de sondaj se vor grupa foarte strîns în jurul valorii adevărate. Dacă facem ca numărul eşantioanelor să crească la infinit, valorile medii de sondaj vor tinde aproape sigur către o limită, care este tocmai media populaţiei originare.

Aşadar, putem vorbi despre o *repartiţie a mediei de sondaj*, avînd valoarea medie m.

Estimarea dispersiei şi abaterii medii pătratice. Datele repartiţiei din tabelul 5.2 pot fi folosite totodată la estimarea dispersiei şi a abaterii medii pătratice, indicatori ce caracterizează gradul de împrăştiere a valorilor populaţiei originare. În acest scop, trebuie să luăm dispersia şi abaterea medie pătratică a eşantionului. În calculul acestor indicatori nu trebuie să folosim abateri de la media adevărată, ci de la media de sondaj ($\bar{x}$), care poate să difere de media adevărată şi care astfel poate afecta valoarea estimaţiei.

Din statistica matematică se ştie că dacă $x_1, x_2, \ldots, x_n$ sînt valorile eşansionului ales şi $\bar{x}$ media lor, atunci sînt motive să preferăm — pentru dispertia populaţiei originare — următoarea estimaţie:

$$\tilde{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 f_i, \tag{5.2}$$

în loc de:

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 f_i. \tag{5.3}$$

Trebuie să observăm însă că dacă n este mare, diferenţa dintre rezultatele pe care le dau expresiile (5.2) şi (5.3) este nesemnificativă şi o putem neglija atît timp cît nu avem de a face cu eşantioane mici. În cazul cercetării noastre aplicative (aplicaţia 1) avem următoarea estimaţie a dispersiei:

$$\tilde{s}^2 = \frac{1}{n-1} \sum_{i=1}^{5} (x_i - \bar{x})^2 f_i = \frac{110\,727\,500}{1\,041} = 1118459{,}59595$$

şi evident a abaterii medii pătratice:

$$\tilde{s} = \sqrt{1118459{,}59595} = 1057{,}5785.$$

Observaţii. Pentru calculul estimaţiei (5.2) cu ajutorul unei maşini de calcul este comodă folosirea următoarei formule:

$$\tilde{s}^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 f_i - \frac{1}{n} \left(\sum_{i=1}^{n} x_i f_i \right)^2 \right]. \tag{5.2'}$$

Dacă într-un caz particular de cercetare oarecare ne este cunoscută valoarea teoretică adevărată m, însă necunoscută valoarea σ, atunci pentru estimarea dispersiei populaţiei originare se foloseşte următoarea formulă

$$s^2 = \frac{1}{n} \sum (x_i - m)^2 f_i, \tag{5.3'}$$

în care la numitor, după cum se vede, apare volumul complet al eşantionului.

Scanned with OKEN Scanner

5.4.3 Eroarea standard a repartiției de sondaj a mediei aritmetice

Am arătat mai sus că media aritmetică de sondaj este o variabilă aleatoare, avînd valoarea medie egală cu m — media populației originare.

Este evident că în cercetările teoretice și aplicative prezintă interes deosebit cunoașterea nu numai a mediei, ci și a abaterii medii pătratice a repartiției mediei de sondaj.

Un rezultat important al cercetărilor de statistică matematică ne arată că dispersia $s_{\bar{x}}^2$ și, respectiv, abaterea medie pătratică a mediei eșantioanelor $s_{\bar{x}}$ depind de abaterea standard (σ) a populației originare, fiind legate de ea prin următoarele relații:

$$s_{\bar{x}}'^2 = \frac{\sigma^2}{n} \tag{5.4}$$

și

$$s_{\bar{x}}' = \frac{\sigma}{\sqrt{n}}. \tag{5.5}$$

Ultima relație definește eroarea standard; în cele ce urmează o vom cita adeseori. Ea este fundamentală în teoria sondajului și determinarea sa are loc fără nici o referire la forma repartiției de frecvență sau la mărimea eșantionului. *Dacă σ ne este cunoscut, relația (5.5) are o aplicabilitate generală.*

În majoritatea cazurilor însă abaterea medie pătratică a populației originare (σ) nu ne este cunoscută și atunci, practic, trebuie să recurgem la estimația ei dată de (5.2).

Așadar, eroarea standard a mediei de sondaj o vom aproxima prin relația:

$$s_{\bar{x}}' = \frac{\sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2 f_i}}{\sqrt{n}}. \tag{5.6}$$

Reținem în mod deosebit atenția cititorului asupra numitorului expresiei (5.5) sau (5.6); după cum ușor se poate observa, odată cu creșterea lui n (volumul eșantionului), eroarea standard a variabilei aleatoare $\bar{x}$ se micșorează direct proporțional cu $\sqrt{n}$. De aici desprindem o concluzie de mare însemnătate: *în cadrul repartiției mediilor de sondaj $\bar{x}$, diferitele lor valori aleatoare se grupează mai strîns în jurul mediei generale m, decît modul în care se plasează — în cadrul repartiției originare — valorile individuale ale diferitelor măsurători în jurul aceleiași medii.*

Tocmai acest rezultat fundamentează teoretic încrederea mai mare pe care o acordăm estimării mediei adevărate cu ajutorul unei medii de sondaj.

În exemplul pe care îl analizăm, avem următoarea valoare a erorii standard a mediei de selecție:

$$s_{\bar{x}}' = \frac{1057{,}5785}{\sqrt{100}} = 105{,}75785.$$

Scanned with OKEN Scanner

5.4.4 Cîteva observaţii cu privire la repartiţia de sondaj a mediei aritmetice

În primul rînd, vom atrage atenţia că repartiţia de sondaj a mediei aritmetice, ca dealtfel şi repartiţiile de sondaj ale multor altor statistici, obţinute în condiţiile unui sondaj simplu, tind, în cazul eşantioanelor mari, către repartiţia normală.

Această observaţie are o justificare teoretică. Se poate arăta că pentru sondaje simple dintr-o populaţie normală, repartiţiile de sondaj ale celor mai multe statistici (a mediei, momentelor etc.) sînt normale, dacă eşantioanele sînt din ce în ce mai mari şi uneori sînt normale şi în cazul eşantioanelor mici.

Unii cercetători au verificat rezultatele teoretice mai sus menţionate în mod experimental asupra unor populaţii care nu sînt normale şi au putut constata că repartiţia originară poate devia destul de mult de la forma normală fără ca prin aceasta normalitatea repartiţiei de sondaj să fie afectată, cu condiţia evidentă ca eşantioanele considerate să fie mari.

Aşadar, dacă asimetria populaţiei originare nu prezintă forme specifice, pronunţate, nu este nevoie să-i cunoaştem forma exactă. Repartiţia de sondaj a unei statistici cum este, de pildă, valoarea medie poate fi reprezentată satisfăcător prin media de sondaj şi eroarea ei standard. Această constatare ne permite să considerăm că un domeniu de variaţie, egal cu de 3 ori eroarea standard, ne oferă limite între care se află, cu mare probabilitate, valoarea adevărată a mediei în populaţia originară.

Am văzut că pentru eşantioane de volum mare — în condiţiile unui sondaj simplu — valorile statisticii în eşantioane pot fi luate drept estimaţii ale valorilor tipice ale populaţiei originare. Mai mult, estimaţiile pot fi folosite la calculul erorilor standard al diferitelor statistici.

Dacă ne referim la exemplul nostru, valoarea medie (m) a populaţiei originare poate fi localizată, cu o probabilitate foarte mare, în următorul interval:

$$\left.\begin{aligned} &\bar{x} - 3 \cdot \frac{s'_{\bar{x}}}{\sqrt{n}} \\ &\bar{x} + 3 \cdot \frac{s'_{\bar{x}}}{\sqrt{n}} \end{aligned}\right\}. \tag{5.7}$$

Considerînd datele noastre concrete, acest interval — consemnat sub o altă formă — este următorul:

$$1415 - 3 \cdot 105{,}75725 < m < 1415 + 3 \cdot 105{,}75725.$$

El trebuie interpretat în termenii teoriei probabilităţilor: cu o probabilitate foarte mare putem afirma că media (m) a populaţiei originare se găseşte în intervalul de variaţie a numărului de muncitori arătat mai sus.

Care anume este nivelul acestei probabilităţi în cazul unor eşantioane mari vom vedea imediat, la expunerea problemei privind construirea diferitelor intervale de încredere pentru aprecierea preciziei estimaţiilor noastre, în ipoteza unei repartiţii teoretice frecvent întîlnite în cercetările economice — repartiţia normală.

Observăm că, în cazul nostru, media populaţiei originare fiind de 1468,80998 ea este localizată în intervalul [1097,72825;1732,27175].

Scanned with OKEN Scanner

5.4.5 Construirea intervalelor de încredere în cazul repartiţiilor normale

S-a observat, desigur, că la estimarea şi localizarea în intervalul arătat a mediei ce caracterizează populaţia originară valoarea mediei de sondaj a fost considerată conform (5.7), împreună cu eroarea standard multiplicată cu un coeficient egal cu 3. Vom preciza acum că acest coeficient poate fi mai mare sau mai mic, în funcţie de nivelul de încredere probabilistică cu care acceptăm limitele intervalului construit; de asemenea, vom introduce expresia de *eroare limită de sondaj* pentru produsul de mai jos $d\frac{s'_{\bar{x}}}{\sqrt{n}}$ din expresia (5.8) de mai jos.

În general, admiţînd ipoteza de normalitate a repartiţiei cercetate şi considerînd un eşantion destul de mare ($n > 30$), limitele intervalului de încredere se consemnează în felul următor:

$$\bar{x} - d\frac{s'_{\bar{x}}}{\sqrt{n}} \; ; \; \bar{x} + d\frac{s'_{\bar{x}}}{\sqrt{n}} \, , \tag{5.8}$$

unde coeficientul d este ales în funcţie de probabilitatea cerută. De pildă, pentru un interval de încredere de 90%, $d = 1{,}645$; pentru un interval de încredere de 95%, $d = 1{,}96$. Cu cît creşte valoarea coeficientului d, gradul de încredere probabilistică acordată intervalului construit creşte şi el; coeficientul d este deci o variabilă aleatoare, de obicei normală normată, avînd valoarea medie egală cu 0 şi dispersia cu 1.

Coeficientul d este ales din tabele speciale. În tabelul 5.3 este dată — în extras — repartiţia teoretică a lui d pentru un număr limitat de valori.

Tabelul 5.3

REPARTIŢIA NORMALĂ NORMATĂ

Nivel de semnificaţie	0,10	0,05	0,02	0,01	0,002	0,001
d	1,645	1,960	2,326	2,576	3,090	3,291

Notă. Tabelul oferă repartiţia normală normată a variabilei aleatoare d, avînd valoarea medie egală cu 0 şi abaterea medie pătratică 1, pentru niveluri de semnificaţie cele mai frecvent folosite la calculul estimaţiilor statistice. De pildă, pentru orice repartiţie normală, probabilitatea evenimentului, ca o abatere de la medie, în valoarea absolută, să depăşească produsul lui 1,96 cu abaterea medie pătratică este egală cu 0,05 sau 5%.

În exemplul nostru am considerat $d = 3$, corespunzător unei probabilităţi a intervalului de încredere de aproximativ 99,8% şi implicit, unui nivel de semnificaţie foarte mic de 0,002 sau 0,02%.

Probabilitatea cu care este garantat un interval de încredere de 0,998 sau 99,8% poate fi scrisă astfel:

$$P\left(\bar{x} - 3 \cdot \frac{s'_{\bar{x}}}{\sqrt{n}} < m < \bar{x} + 3 \cdot \frac{s'_{\bar{x}}}{\sqrt{n}}\right) = 0{,}998 \text{ sau } 99{,}8\%.$$

Scanned with OKEN Scanner

Un interval mai restrîns, garantat cu o probabilitate de aproximativ 0,95 sau 95% este următorul:

$$P\left(\bar{x} - 1{,}96\frac{s'_{\bar{x}}}{\sqrt{n}} < m < \bar{x} + 1{,}96\frac{s'_{\bar{x}}}{\sqrt{n}}\right) = 0{,}95 \text{ sau } 95.$$

În exemplul nostru concret putem scrie:

$$P(1415 - 1{,}9 \cdot 105{,}75725 < m < 1415 + 1{,}96 \cdot 10575725 = 95\%.$$

O atenție deosebită trebuie să acordăm interpretării acestei probabilități. Practic, aceasta înseamnă că în aproximativ 95 din 100 eșantioane aleatoare de mărimea celui folosit de noi, putem considera că media estimată a populației generale se va afla între limitele numărului de muncitori egale cu 1207,71579 și 1622,28421.

Or, avînd la dispoziție media populației generale de 1468,80998 putem constata că ea se localizează perfect și în acest interval de încredere mai restrîns [1207,71579; 1622,28421] garantat cu o probabilitate de numai 95%.

Conceptul intervalului de încredere se aplică pe o scară mai largă decît cel al erorii standard. Într-adevăr, eroarea standard poate fi interpretată mai lesne numai în acele cazuri în care repartiția variabilei supusă studiului este apropiată de o repartiție normală. Cît privește intervalele de încredere, ele pot fi găsite pentru o clasă mult mai largă de repartiții; ne vom întîlni cu asemenea exemple, mai tîrziu, cînd vom analiza informația statistică oferită de eșantioane mici, cu ajutorul repartiției lui Student.

5.5. CRITERII ȘI TESTE DE APRECIERE A SEMNIFICAȚIEI STATISTICE A VALORILOR TIPICE DE SONDAJ

La sfîrșitul paragrafului precedent am introdus — alături de conceptele de nivel de încredere și nivel de semnificație — conceptul de variabilă aleatoare normală normată, simbolizată prin coeficientul d, ale cărui valori, multiplicate cu eroarea standard a mediei de sondaj $\frac{s'_{\bar{x}}}{\sqrt{n}}$, ne-au servit drept instrument de măsură a preciziei cu care estimăm, pe un anumit interval, valoarea medie a populației originare.

5.5.1 Ipoteza diferenței nule

Nivelul de încredere în oricare din indicatorii statistici de măsură obținuți dintr-un eșantion — media, mediana, moment centrat etc. — ca estimații ale parametrilor respectivi din populația originară, depinde, în primul rînd, de *stabilitatea* acestor indicatori de sondaj.

Cum stabilitatea poate fi însă pusă ușor în evidență cu ajutorul conceptului *împrăștierii* (variabilității), în statistica matematică s-a născut ideea ca tocmai indicatorii împrăștierii și, în special, abaterea medie pătratică, să fie folosiți drept *criterii* de apreciere a *semnificației* mărimilor de sondaj obținute.

Scanned with OKEN Scanner

Pe de altă parte — și asupra acestui aspect al problemei am atras atenția încă în 5.1 — mărimile obținute pe bază de sondaj vor estima cu atît mai precis valorile adevărate cu cît volumul eșantionului — îndeplinind condiția de a fi aleator și reprezentativ — va fi mai mare.

Rezultă că, în afară de indicatorii variabilității, volumul eșantionului n este implicat și el în formularea unui criteriu oarecare de semnificație statistică.

Un criteriu de apreciere a semnificației statistice poate căpăta forme diferite de expresie, în funcție de *natura* problemei ce trebuie rezolvată. În principiu însă, firul raționamentului logic folosit la formularea unui criteriu dat de semnificație statistică rămîne același.

În general, aplicarea diferitelor criterii pentru aprecierea semnificației statistice a indicatorilor de sondaj poate fi fundamentată din punct de vedere teoretic însă cea mai elocventă probă a eficienței lor — cel puțin atunci cînd criteriile respective sînt folosite în munca de cercetare — rezidă în folosul practic pe care-l oferă cercetătorului.

Vom ilustra modul de construcție a unui criteriu de semnificație ca și raționamentul folosit cu acest prilej.

În primul rînd se stabilește o ipoteză de lucru — de obicei ea are un caracter general — denumită în statistică *ipoteză a diferenței nule* [4]. De la o problemă concretă de cercetare la alta, această ipoteză capătă un sens specific problemei respective. În esență, ea postulează faptul că *nu există nici o diferență* semnificativă între indicatorii de sondaj și cei ai populației originare sau între indicatorii de sondaj de la un eșantion aleator la altul. Este important însă să precizăm că această ipoteză de lucru pornește de la o anumită concepție a cercetătorului pe care el o formulează fie *apriori*, deci înainte de a fi întreprins orice fel de măsurători, observații sau experimente, fie — și aceasta în orice caz — înainte de a trece la analiza de detaliu a datelor obținute.

De pildă, cel ce cercetează, statistic, prin experimentări repetate, frecvența de apariție a uneia din fețele zarului poate porni de la ideea că zarul este perfect omogen și simetric, astfel încît probabilitatea teoretică de apariție a oricăreia din fețele zarului este egală cu 1/6 și apoi să formuleze, pe această bază, ipoteza diferenței nule între frecvențele statistice și cele teoretice. Trecînd la efectuarea experimentului, cercetătorul nostru își va apăra, desigur, punctul său de vedere propriu, chiar dacă frecvențele se vor deosebi întrucîtva de cele teoretice, testînd, în acest scop, pe baza unui procedeu corespunzător, ipoteza diferenței nule. Dacă însă, în procesul experimentării, frecvențele observate se vor dovedi a fi total aberante, neașteptat de îndepărtate de cele teoretice, atunci cercetătorul nostru va trebui să decidă între a abandona teoria sa inițială sau a o reformula esențial.

Exemple analoage putem da și din alte domenii de cercetare. Este cunoscută astfel teoria pe care o acceptă, în general, cercetătorii demografi, cu privire la proporțiile de aproximativ 0,51 și 0,49 ale născuților-vii în favoarea celor de sex masculin sau teoria că sexul fiecărui copil nu depinde de sexul oricărui alt copil născut anterior de aceeași mamă ș.a.m.d.

În următoarele două paragrafe ne vom ocupa de cîteva criterii de semnificație simple, în cazul unor repartiții normale.

[4] În literatura de specialitate se întîlnesc și expresiile: ipoteză nulă sau ipoteză de nul. Nouă ni se pare că expresia folosită în acest context este cea mai potrivită.

Scanned with OKEN Scanner

5.5.2 Criterii de semnificație bazate pe repartiții normale

Compararea unei medii de sondaj cu media cunoscută a populației originare

Aplicația 2. Să presupunem că dintr-o îndelungată activitate de observare statistică, prin măsurarea greutății curcanilor de o anumită rasă și vîrstă în toate unitățile avicole de stat ale unui județ oarecare, ne este bine cunoscut faptul că repartiția greutății este normală. Desigur ne mai sînt cunoscute și valorile principalilor parametri ai populației generale — media și abaterea medie pătratică.

Fie aceste valori: $m = 15$ kg și $\sigma = 4$ kg.

Să presupunem, mai departe, că în atenția noastră se află una din întreprinderile avicole de stat ale județului respectiv — să zicem întreprinderea A, — în cadrul căreia, prin sondaj aleator, alegem 36 curcani și constatăm că greutatea medie a unui curcan în acest eșantion este de 13 kg.

Această constatare ne sugerează următoarea întrebare: este oare semnificativă diferența dintre media de sondaj constatată și media greutății în populația generală de 15 kg?

După cum ne este cunoscut, eroarea standard a mediei de sondaj o obținem determinînd raportul $\frac{\sigma}{\sqrt{n}}$ care, în cazul nostru, este $4/\sqrt{36} = 4/6$.

În cadrul cercetării întreprinse, să considerăm acum ipoteza diferenței nule — cu un nivel de semnificație de 0,05 sau 5%, — admițînd astfel, ipotetic, că eșantionul nostru de 36 curcani în întreprinderea A, este reprezentativ pentru întreaga populație de curcani a județului și, în consecință, diferența dintre media de sondaj și media populației generale (în cazul dat, 2 kg) este o variabilă aleatoare ce se supune unei repartiții normale, avînd media egală cu zero.

Diferența de sondaj observată, exprimată în erori standard ale mediei de sondaj este egală cu $d = 2 : 4/6$, adică 3 unități, valoare căreia îi corespunde un nivel de semnificație de numai 0,002 sau 0,2%.

S-a aplicat formula:

$$d = \frac{m - \bar{x}}{\frac{\sigma}{\sqrt{n}}} = \frac{15 - 13}{4/6} = 3.$$

Este clar că situația examinată nu poate fi considerată drept foarte probabilă. Într-adevăr, noi știm că în cazul unei repartiții normale, 5% din valorile observate se găsesc în acea parte a curbei de repartiție care se află de o parte și de alta a valorii medii la o distanță mai mare decît o unitate a abaterii medii pătratice, multiplicată cu 1,96.

Or, dacă raportul dintre o diferență oarecare — rezultată din datele de observație — și eroarea standard este *mai mare* de 1,96, atunci, în cazul unui nivel de semnificație de 5%, această diferență nu poate fi atribuită fluctuațiilor de sondaj, ci trebuie considerată drept *semnificativă* și în consecință vom trage concluzia că ipoteza diferenței nule trebuie respinsă.

Așadar, din punctul de vedere al aprecierii generale economico-gospodărești, se poate conchide că la întreprinderea A există *cauze sistematice* care afectează negativ greutatea normală a curcanilor (subnutriție, neîngrijire rațională etc.) și care, odată înlăturate, vor deschide perspective mai bune

Scanned with OKEN Scanner

și în cadrul acestei ferme pentru asigurarea parametrilor de greutate caracteristici pentru întreaga regiune.

În general, în practica cercetărilor pe bază de sondaj noi lucrăm cu un eșantion format din n elemente, cu media caracteristicii cercetate $\bar{x}$ și abaterea medie pătratică de sondaj s și dorim să comparăm media de sondaj cu media m, cunoscută, a populației originare, admițînd ipoteza diferenței nule; dacă nu cunoaștem nici abaterea medie pătratică a populației originare o putem înlocui cu valoarea ei de estimație s. După aceea determinăm mărimea d, făcînd raportul:

$$d = \frac{m - \bar{x}}{s/\sqrt{n}}.$$

Considerînd că d este o variabilă aleatoare ce se supune unei legi normale de repartiție, avînd valoarea medie egală cu 0 și abaterea medie pătratică 1, putem determina nivelul de semnificație pe care l-am atins, pe baza tabelului special de repartiție a probabilităților care ne este cunoscut.[5]

Compararea mediilor a două eșantioane mari. Vom examina acum situația în care sînt extrase două eșantioane aleatoare din aceeași populație, ale căror medii sînt diferite. În mod firesc se pune problema de a ști dacă diferența constatată poate fi atribuită exclusiv fluctuațiilor de sondaj.

Dacă cele două eșantioane sînt independente și provin din aceeași populație, în condițiile sondajului simplu, atunci eroarea standard a diferenței mediilor notată cu $s'_{\bar{x}_1-\bar{x}_2}$ este dată de relația:

$$s'_{\bar{x}_1-\bar{x}_2} = \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}. \tag{5.9}$$

Dacă diferența observată depășește triplul valorii lui $s'_{\bar{x}_1-\bar{x}_2}$ dată de relația (5.9), este evident că ea poate fi cu greu atribuită fluctuațiilor de sondaj.

Deseori, în practică, valoarea lui σ nu ne este cunoscută apriori și deci trebuie înlocuită cu o valoare observată. În mod firesc ajungem la ideea de a considera această valoare ca fiind egală cu abaterea medie pătratică în cele două eșantioane considerate *împreună*. Totuși, dacă abaterile medii pătratice ale celor două eșantioane diferă *semnificativ* nu putem conchide că ambele eșantioane au fost extrase din una și aceeași populație. Rezultă că, dacă două eșantioane sînt extrase din populații diferite, fiecare avînd dispersiile s_1^2 și s_2^2, eroarea standard a diferenței mediilor lor va fi dată de relația:

$$s'_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}. \tag{5.10}$$

Tocmai această relație este folosită, în cazul general, drept *test* al semnificației diferenței dintre două medii de sondaj. Ținînd seama că eroarea standard a mediei de sondaj, $s'_{\bar{x}}$, depinde numai de abaterea medie pătratică a populației originare și nu de media ei, problema dacă cele două populații din care au fost extrase eșantioanele diferă semnificativ o putem elucida comparînd doar valorile lor medii, *independent de eventualele deosebiri în gradul lor de variabilitate.*

[5] Vezi tabelul 5.3.

Scanned with OKEN Scanner

Această comparație o putem face în mod analog cu cel oferit de relația (5.9), introducînd și în acest caz estimațiile corespunzătoare ale lui σ_1^2 și σ_2^2. Vom determina, astfel, mărimea d, făcînd raportul:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{x_2^2}{n_2}}}. \tag{5.11}$$

După cum se vede din relația (5.11), mărimea variabilă d este pozitivă cînd $\bar{x}_1 > \bar{x}_2$ și negativă cînd $\bar{x}_1 < \bar{x}_2$. Dacă valoarea absolută a lui d este mai mare decît 1,96, atunci la un nivel de semnificație de 5% putem considera că *există* o diferență *semnificativă* între mediile necunoscute m_1 și m_2 ale populațiilor originare. Astfel, dacă $d = -2{,}20$, la nivelul ales al semnificației de 5%, acest rezultat este semnificativ.

Este, desigur, util și calculul limitelor intervalului de încredere pentru diferența $m_1 - m_2$ dintre două medii ale populațiilor originare. Ele vor fi egale cu:

$$(\bar{x}_1 - \bar{x}_2) - d \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{și} \quad (\bar{x}_1 - \bar{x}_2) + d \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{5.12}$$

în care, ca de obicei, mărimea lui d se ia din tabele speciale, în conformitate cu nivelul probabilității cerute.

Este de reținut următoarea observație importantă.

În ipoteza în care două eșantioane independente sînt extrase din aceeași populație și, în loc să comparăm media unuia cu media celuilalt, vom compara, de pildă, media primului $\bar{x}_1$ cu media $\bar{x}_0$ a ambelor eșantioane considerate împreună, folosirea relațiilor (5.9) și (5.10) nu este justificată, întrucît erorile ce-și exercită influența asupra mediei unui eșantion sînt *corelate* cu cele ce-și exercită influența asupra mediei $\bar{x}_0$ a celor două eșantioane considerate împreună.

Se poate arăta că, în acest caz, eroarea standard a diferenței celor două medii, $s'_{\bar{x}_1 - \bar{x}_2}$, este următoarea:

$$s'_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma^2 \frac{n_2}{n_1(n_1 + n_2)}}. \tag{5.13}$$

Pentru a ilustra aplicarea formulelor prezentate mai sus, vom considera următoarea aplicație.

Aplicația 3. Pe teritoriul unui județ există în proprietate personală un număr de 3500 vaci. În condițiile riguroase ale unui sondaj simplu, se extrag două eșantioane, primul avînd un volum de 900 vaci, iar al doilea de 400 vaci.

Eșantioanele s-au ales independent și în fiecare din ele s-a cercetat cantitatea de lapte muls în decurs de un an pe fiecare vacă.

Eșantioanele și informațiile oferite de ele sînt următoarele

Primul eșantion	*Al doilea eșantion*
1) $n_1 = 900$ vaci	$n_2 = 400$
2) $x_1 = 3042$ l (cantitatea anuală de lapte muls pe o vacă)	$x_2 = 3000$ l (Idem)
3) $s_1 = 300$ l	$s_2 = 300$ l

Scanned with OKEN Scanner

Ne interesează să decidem dacă diferența dintre cele două medii $\bar{x}_1 - \bar{x}_2 = 3042 - 3000 = +42$ este *semnificativă*. Vom considera suficient ca această decizie să o putem lua la un nivel de semnificație de 5%.

Se pune astfel problema de a decide dacă, în speță, ipoteza diferenței nule poate fi acceptată sau trebuie respinsă, adică dacă cele două eșantioane au fost extrase din aceeași populație originară sau din două populații diferite sub aspectul caracteristicii luate în studiu.

Avînd în vedere datele oferite de cele două eșantioane, calculul indicatorilor respectivi și raționamentele care le însoțesc se vor desfășura după cum urmează:

1. Se calculează erorile standard ale celor două medii:

$$s'_{\bar{x}_1} = \sqrt{\frac{s_1^2}{n_1}} = \sqrt{\frac{300^2}{900}} = \frac{300}{30} = 10 \text{ l.}$$

$$s'_{\bar{x}_2} = \sqrt{\frac{s_2^2}{n_2}} = \sqrt{\frac{300^2}{400}} = \frac{300}{20} = 15 \text{ l}$$

2. Se calculează eroarea standard a diferenței dintre medii:

$$s'_{\bar{x}_1 - \bar{x}_2} = \sqrt{s'^2_{\bar{x}_1} + s'^2_{\bar{x}_2}} = \sqrt{10^2 + 15^2} \simeq 18 \text{ l.}$$

3. Se calculează valoarea lui d, potrivit relației (5.11)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s'_{\bar{x}_1 - \bar{x}_2}} = \frac{3042 - 3000}{18} = \frac{42}{18} \simeq 2{,}33.$$

4. Se face evaluarea lui d în termenii teoriei probabilităților pentru verificarea ipotezei diferenței nule dintre cele două medii.

În procesul acestei evaluări sînt de subliniat două momente importante:

a) *Se determină probabilitatea ca o valoare a lui d, egală cu sau mai mare decît valoarea obținută, să apară pe bază de variații întîmplătoare de sondaj*, în alegerea eșantioanelor provenite din aceeași populație. Această probabilitate se obține, după cum știm, dintr-un tabel special.

În aplicația noastră am obținut pentru d o valoare aproximativă $d \simeq 2{,}33$. Din tabelul amintit al probabilităților se vede că există doar o șansă de 2% ca o valoare a lui d — pe bază de variații aleatoare la alegerea eșantioanelor — să fie egală *cu* sau *mai mare* decît cea pe care am obținut-o.

b) Cunoscînd această probabilitate tragem, în încheiere, o concluzie asupra rațiunii de a fi a ipotezei formulate — ipoteza diferenței nule. Dacă probabilitatea este foarte mică, ca în cazul nostru, ipoteza se respinge ca fiind improbabilă, iar diferența obținută se socotește ca fiind semnificativă. Aceasta înseamnă că cele două eșantioane reprezintă, probabil, două populații diferite. În cazul nostru, în care am acceptat un nivel de semnificație de 5%, pentru admiterea ipotezei formulate ar fi trebuit să obținem, pe baza celor două eșantioane, $d = 1{,}96$. Așadar, diferența celor două medii trebuie considerată ca fiind semnificativă.

În termeni de limite de încredere se poate afirma — avînd în vedere nivelul de semnificație ales — că, dacă s-ar respinge în mod continuu ipoteze

ale diferenței nule, la *nivelul* de încredere de 5%, în 5% din cazuri ele s-ar fi respins în mod *eronat*.

Este de notat că în cadrul celor mai multe activități de cercetare se admit următoarele criterii privind acceptarea sau respingerea ipotezei diferenței nule:

— se *acceptă* ipoteza și se consideră deci diferențele *nesemnificative*, cînd d este mai mic decît 1,96 (nivel de semnificație de 5%);

— se *respinge* ipoteza și se consideră astfel diferențele *semnificative*, cînd d este mai mare de 2,58 (nivel de semnificație de 1%);

— *ipoteza nu se acceptă, dar nici nu se respinge* — considerîndu-se diferența ca fiind de semnificație îndoielnică — cînd d se situează între 1,96 și 2,58. În acest caz s-a încadrat aplicația noastră.

După cum se vede, o ipoteză a diferenței nule nu poate fi niciodată ori respinsă, ori acceptată cu o siguranță totală, ea se poate dovedi a fi foarte improbabilă, foarte probabilă sau intermediară între aceste extreme.

Cele trei nivele de semnificație, aplicate adeseori în practică — 5%, 1% și 0,2% — cărora le corespund valorile respective ale variabilei $d = 1{,}96$, $d = 2{,}58$ și $d = 3$, fiecare implicînd un risc de respingere pe nedrept din ce în ce mai mic, sînt destul de arbitrare, astfel încît cercetătorul trebuie să fie foarte prudent în alegerea unuia dintre ele. Totul depinde de *natura* problemei și de *riscul* implicat în acceptarea pe nedrept sau respingerea eronată a ipotezei respective. Dacă, de pildă, o acceptare falsă sau o respingere falsă ar implica riscul pierderii vieții, vătămării integrității corporale, pierderii averii sau onoarei, este clar că în aceste cazuri trebuie folosite nivelurile de semnificații cele mai riguroase, reducîndu-se la minimum riscul de a se respinge o ipoteză corectă sau de a se accepta o ipoteză greșită.

În cazul nostru, $d = 2{,}33$, deci $d > 1{,}96$, diferența fiind semnificativă la un nivel de semnificație de 0,05; însă $d < 2{,}85$ — diferența nesemnificativă la un nivel de semnificație de 0,01 și, cu atît mai mult, dacă $d < 3$ diferența este nesemnificativă cu probabilitatea 0,2%.

Aceste diferențe, după cum am arătat mai sus, nu reprezintă o contradicție. În primul caz, dacă admitem că diferențele nu sînt semnificative, putem greși în 5% din cazuri; dacă sîntem mai pretențioși și nu dorim să facem o afirmație eronată decît cu probabilitatea 0,01, atunci spunem că diferențele sînt semnificative.

5.6. EROAREA STANDARD A MEDIANEI

După cum se știe, în cazul variabilelor strict discontinue, mediana este o formă de medie de un sens nesigur și de aceea utilizarea ei în asemenea caz trebuie evitată [6]. Nu același lucru se poate spune și despre cazul unor serii de observații cu variații *continue*, formate pe bază de eșantioane mari. În acest ultim caz, mediana poate să fie uneori preferabilă mediei, datorită faptului că este mai puțin afectată de valorile anormal de mari sau anormal

[6] Vezi explicațiile date în 4.4.2.

de mici ale caracteristicii variabile. Dacă observaţiile prezintă asemenea *valori extreme mari*, mediana va fi *mai stabilă* şi mai puţin afectată de fluctuaţiile aleatoare de sondaj decît media aritmetică.

Pe de altă parte, dacă ţinem seama că mediana este o formă concretă şi naturală de medie şi că elementul median al populaţiei, într-un sistem de măsurare, rămîne median şi într-un alt sistem de măsurare, cu condiţia evidentă ca noul sistem să păstreze toate elementele populaţiei, în aceeaşi ordine de mărime, devine clar că în diferite cercetări ştiinţifice este foarte utilă *determinarea erorii standard a medianei* care diferă, ca mărime, de eroarea standard a mediei aritmetice; ea poate fi apoi folosită la construirea intervalului respectiv de încredere prin care se va estima valoarea medianei în populaţia originară.

Dacă eroarea standard a medianei o notăm cu s'_{Me} atunci se poate arăta că ea este definită de următoarea relaţie:

$$s'_{Me} = \frac{\sigma}{\sqrt{n}} \cdot 1{,}25331. \qquad (5.14)$$

În general, acest raport se menţine şi în cazul unor repartiţii care nu diferă mult de forma normală.

Am arătat mai sus că în anumite cazuri excepţionale, mediana ar putea să fie mai puţin influenţată de erorile de sondaj şi astfel să fie mai stabilă. Este vorba de repartiţii care se întîlnesc în cercetările de statistică economică; ele au forme cu vîrfuri ascuţite şi ramuri de lungimi relativ mari. Dacă, în cazul acestor repartiţii, stabilitatea mai mare a medianei este suficient de evidentă, atunci această valoare poate fi cea mai bună valoare tipică ce se poate folosi în cercetarea respectivă, compensîndu-se astfel dezavantajele ei în alte privinţe.

Aplicaţia 4.[7] Prin sondaj simplu s-a ales un eşantion format din 100 cooperative agricole de producţie în cadrul cărora s-a cercetat nivelul recoltei la produsul A. Prelucrîndu-se în mod corespunzător informaţiile obţinute, s-a putut obţine repartiţia de sondaj din tabelul de mai jos.

Tabelul 5.4

Recolta la ha (q) (intervale de grupare)	Mijlocul intervalului	
6,4— 8,8	7,6	4
8,8—11,2	10,0	21
11,2—13,6	12,4	30
13,6—16,0	14,8	17
16,0—18,4	17,2	14
18,4—20,8	19,6	7
20,8—23,2	22,0	3
23,2—25,6	26,8	—
25,6—28,0	29,2	1
28,0—30,4	31,6	1
30,4—32,8	31,6	3

[7] Aplicaţia a fost construită după modelul din lucrarea: M. P. ALTUNIN. *Prakticeskoie posobie po obşcei i selshohazeaistvenoi statistike*, Moscova, 1947, p. 218.

Scanned with OKEN Scanner

Vom calcula media şi mediana acestei repartiţii, ca şi erorile lor standard, ţinînd seama de relaţiile ce le definesc:

Media: $= 14{,}277$

Eroare standard a mediei

$$s'_{\bar{x}} = \frac{\sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}}}{\sqrt{100}} = 0{,}465952$$

Mediana:

$$\text{Me} = V_0 + i\left(\frac{\frac{n}{2} - F_{i-1}}{f_i}\right) = 11{,}2 + 2{,}4\left(\frac{50-25}{30}\right) = 13{,}2$$

în care: V_0 este limita inferioară a intervalului median;
i — mărimea intervalului;
n — volumul eşantionului;
F_{i-1} — frecvenţa cumulată pînă la intervalul median;
f_i — frecvenţa de grupă în intervalul median.

Eroarea standard a medianei. În calculul acestui indicator vom folosi relaţia (5.14):

$$s'_{Me} = \frac{\sqrt{s'^2_{\bar{x}}}}{\sqrt{100}} \cdot 1{,}25331 = 0{,}583982.$$

Pentru a estima valoarea medianei în populaţia originară putem construi următorul interval de încredere, cu un nivel de semnificaţie de 0,2%:
$13{,}2 - 3 \cdot 0{,}583982$ şi $13{,}2 + 3 \cdot 0{,}583982$.

Cu alte cuvinte, admiţînd ipoteza normalităţii, cu o probabilitate de aproximativ 0,998 se poate afirma că în populaţia originară mediana va fi cuprinsă între limitele de 11,448054 şi 14,951946 q/ha.

5.7. ERORILE STANDARD ALE MOMENTELOR DE ORDIN SUPERIOR ŞI ALE ALTOR VALORI TIPICE DE SONDAJ

Trecem la studiul erorilor standard ale altor importante valori tipice ale repartiţiilor de sondaj.

5.7.1 Erorile standard ale momentelor convenţionale [8] şi centrate

Cea mai generală expresie — o dăm fără demonstraţie —, care defineşte *eroarea standard a unui moment convenţional*, este următoarea:

$$s'_{\lambda_k} = \sqrt{\frac{\lambda_{2,k} - \lambda_k^2}{n}}, \qquad (5.15)$$

în care λ_k reprezintă momentul convenţional de ordinul k.

[8] Amintim că prin momente *convenţionale* înţelegem momente calculate în raport cu o valoare oarecare a caracteristicii studiate, arbitrar aleasă (vezi 4.4.2.)

Scanned with OKEN Scanner

Aplicația 5. Ne propunem să calculăm erorile standard ale momentelor convenționale de ordinul doi și trei, caracterizînd eșantionul aleator de 100 c.a.p. din aplicația precedentă și apoi să construim intervalele respective de încredere.

În acest scop am calculat valorile primelor șase momente convenționale, considerînd abaterile ξ_s de diferite ordine de la punctul fix $A = 17{,}2$ (mijlocul intervalului al cincilea), măsurate în lungimi de interval ($i = 2{,}4\ q$), adică:

$$\xi_s^{(i)} = \frac{x - A}{i}.$$

Rezultatul calculelor este următorul:

$$\lambda_1 = -1{,}22; \quad \lambda_2 = 5{,}22; \quad \lambda_3 = -4{,}28; \quad \lambda_4 = 67{,}50$$

$$\lambda_5 = 96{,}28; \quad \lambda_6 = 1468{,}62.$$

Erorile standard ale momentelor convenționale de ordinul 2 și 3, — aplicînd formula (5.15) — sînt următoarele:

$$s'_{\lambda_2} = \sqrt{\frac{\lambda_4 - \lambda_2^2}{n}} = \sqrt{\frac{67{,}50 - 5{,}22^2}{100}} \simeq 0{,}6344 \text{ intervale}$$

$$s'_{\lambda_3} = \sqrt{\frac{\lambda_6 - \lambda_3^2}{n}} = \sqrt{\frac{1468{,}62 - (-4{,}28^2)}{100}} \simeq 3{,}8082 \text{ intervale.}$$

Admițînd ipoteza că legea de repartiție a recoltei la produsul A în cadru populației originare este normală cu o probabilitate de 0,998, putem afirma că valoarea momentului de ordinul doi va fi cuprinsă în intervalul $5{,}22 \pm 3 \cdot 0{,}6344$. Cu aceeași probabilitate putem afirma că valoarea momentului trei, ce caracterizează populația originară, va fi cuprinsă în intervalul $-4{,}28 \pm 3 \cdot 3{,}8082$.

În practica cercetărilor statistice, de cele mai multe ori, recurgem la *momente centrate* de sondaj, pentru a caracteriza diferite repartiții luate în studiu, momente în care — după cum ne este cunoscut — se consideră abaterile în raport cu media de sondaj. Or, această medie este ea însăși supusă fluctuațiilor aleatoare de sondaj și de aceea erorile standard ale momentelor centrate nu vor fi, în general, aceleași cu cele obținute în cazul momentelor convenționale.

Se poate arăta că *eroarea standard a momentului centrat de ordinul k* — îl vom nota cu μ_k — este definită de următoarea expresie:

$$S'_{\mu_k} = \sqrt{\frac{\mu_{2k} - \mu_k^2 + k^2\mu_2 \cdot \mu_{k-1} - 2k\mu_{k-1} \cdot \mu_{k+1}}{n}} \qquad (5.16)$$

Pentru a putea face cîteva aplicații particulare ale expresiei (5.16) vom recurge din nou la eșantionul celor 100 de c.a.p., calculînd de data aceasta valorile primelor șase momente centrate. Calculul ne-a condus la următoarele rezultate (în unități de măsură obișnuite):

$$\mu_1 = 0; \quad \mu_2 = 21{,}4940; \quad \mu_3 = -154; \quad \mu_4 = 2873{,}7256;$$

$$\mu_5 = 42014{,}16; \quad \mu_6 = 707163{,}39.$$

Scanned with OKEN Scanner

Aplicația 6. Considerînd valorile de mai sus, ne punem problema de a determina — aplicînd formula (5.16) — erorile standard ale dispersiei și ale momentelor centrate de ordinul trei și patru.

1°. *Eroarea standard a dispersiei.* Dacă în expresia (5.16) punem $k = 2$, obținem, dat fiind că $\mu_1 = 0$:

$$s'_{\mu_2} = s_{s^2} = \sqrt{\frac{\mu_4 - \mu_2^2}{n}}, \tag{5.17}$$

introducînd datele cunoscute privind valoarea momentelor centrate, exprimate în unități de măsură obișnuite, ajungem la următorul rezultat:

$$s'_{\mu_2} = s'_{s^2} = \sqrt{\frac{2873{,}7256 - 21{,}4940^2}{100}} = \frac{49{,}1094}{\sqrt{100}} = 4{,}91094.$$

2°. *Eroarea standard a momentelor centrate de ordinul trei și patru.* Dacă în ecuația (5.16) luăm $k = 3$, obținem următoarea expresie pentru eroarea standard a momentului centrat, de ordinul trei:

$$s'_{\mu_3} = \sqrt{\frac{\mu_6 - \mu_3^2 + 9\mu_2\mu_{3-1} - 6\mu_2\mu_4}{n}}. \tag{5.18}$$

Sau, introducînd datele:

$$s'_{\mu_3} = \sqrt{\frac{707163{,}39 - (-154{,}7403)^2 + 9 \cdot 21{,}494^2 - 6 \cdot 21{,}494 \cdot 2873{,}7256}{100}} \simeq 63{,}4268.$$

Raționînd analog, dacă în ecuația (5.16) luăm $k = 4$, obținem — pentru eroarea standard a momentului centrat de ordinul patru — următoarea expresie:

$$s'_{\mu_4} = \sqrt{\frac{\mu_8 - \mu_4^2 + 16\mu_2\mu_3 - 8\mu_3\mu_5}{n}}. \tag{5.19}$$

Calculul valorii concrete a acestei erori standard — considerînd datele repartiției de sondaj din tabelul 5.3. — îl lăsăm, ca exercițiu, pe seama cititorului.

Observație. Din calculul erorii standard a momentului centrat de ordinul trei și din cel al erorii standard pe care cititorul interesat îl va face pentru momentul centrat de ordinul patru se poate ușor observa că erorile standard ale momentelor de ordin superior sînt foarte mari și acest lucru este firesc dacă ținem seama de faptul că momentele cresc rapid, iar eroarea standard a momentului de ordinul k depinde de momentul de ordinul $2k$. De aceea, în cele mai multe probleme practice, erorile standard pentru momentele de ordin superior în eșantioanele de volum mare se calculează numai pentru primele patru momente.

5.7.2 Eroarea standard a diferenței dintre două abateri medii pătratice de sondaj

Analog metodologiei folosite la compararea mediilor a două eșantioane — în ipoteza că ele sînt independente și provin din aceeași populație — putem compara și abaterile medii pătratice din eșantioanele respective.

Scanned with OKEN Scanner

Eroarea standard a diferenței dintre abaterile medii pătratice este definită de următoarea expresie:

$$s'_{(s_1-s_2)} = \sqrt{\frac{\mu_4 - \mu_2^2}{4\mu_2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \tag{5.20}$$

unde n_1 și n_2 sînt volumele celor două eșantioane.

Dacă populația originară urmează o lege de repartiție normală, eroarea standard a diferenței dintre abaterile medii pătratice este dată de relația

$$s'_{(s_1-s_2)} = \sqrt{\frac{\sigma^2}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \tag{5.21}$$

Nu este exclus ca cele două eșantioane să fie extrase din populații diferite, caracterizate prin parametrii $\mu_{2(1)}$, $\mu_{4(1)}$ și respectiv $\mu_{2(2)}$; $\mu_{4(2)}$; în acest caz eroarea standard a diferenței dintre abaterile medii pătratice este definită astfel:

$$s'_{(s_1-s_2)} = \sqrt{\frac{\mu_{4(1)} - \mu_{2(1)}^2}{4\mu_{2(1)} \cdot n_1} + \frac{\mu_{4(2)} - \mu_{2(2)}^2}{4\mu_{2(2)}^2 \cdot n_2}} \tag{5.22}$$

sau, dacă populația este normală:

$$s'_{(s_1-s_2)} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma^2}{2n_2}}. \tag{5.23}$$

În sfîrșit, mai putem compara abaterea medie pătratică a unui eșantion cu abaterea medie pătratică a celor două eșantioane considerate laolaltă. Pe acest caz considerăm ipoteza de normalitate — putem folosi eroarea standard a diferenței de forma:

$$s'_{(s_0-s_1)} = \sqrt{\frac{\sigma^2}{2}\left(\frac{n_2}{n_1(n_2 + n_1)}\right)}. \tag{5.24}$$

Rezultatele (5.20) — (5.24) pot fi folosite la testarea caracterului semnificativ sau nesemnificativ al diferenței dintre abaterile medii pătratice.

5.7.3 Eroarea standard a coeficientului de variație

Se știe că unul din indicatorii clasici de măsurare a variabilității unei repartiții — coeficientul de variație V — este definit prin relația:

$$V = \frac{100 \cdot \sigma}{\bar{x}} = \frac{\sqrt{\mu_2}}{\bar{x}} \cdot 100. \tag{5.25}$$

De pildă, valoarea lui concretă în cazul repartiției de sondaj din tabelul 3 — aplicația 4 — este:

$$V = \frac{100 \cdot 0{,}46357}{14{,}272} = 0{,}3242 \text{ sau } 32{,}48\%.$$

Scanned with OKEN Scanner

Eroarea standard a coeficientului de variaţie este definită de relaţia:

$$s'_v = \frac{V}{\sqrt{2n}} \sqrt{1 + \frac{2V^2}{10^4}}. \qquad (5.26)$$

În majoritatea cazurilor practice al doilea factor diferă puţin de 1, astfel încît $\frac{V}{\sqrt{2n}}$ va da, de regulă, un rezultat destul de precis.

În cazul repartiţiei noastre avem:

$$s'_v = \frac{0{,}3248}{\sqrt{2 \cdot 100}} \sqrt{1 + \frac{2 \cdot 0{,}3248^2}{10\,000}} = 0{,}02290.$$

Interpretînd acest rezultat în termeni de interval de încredere putem afirma că este foarte puţin probabil ca valoarea coeficientului de variaţie în populaţia originară să se găsească în interiorul intervalului de $0{,}3248 \pm \pm 3 \cdot 0{,}02290$, adică $32{,}48\% \pm 6{,}87\%$.

5.8 SONDAJ TIPIC (STRATIFICAT)

În cazul în care se ştie că populaţiile originare sînt puternic neomogene sub aspectul caracteristicii luate în studiu, cercetarea lor pe bază de sondaj trebuie precedată de o operaţie de grupare — stratificarea —, pentru a se obţine un număr de s grupe tipice, cît mai omogene cu putinţă. Fiecare din aceste grupe contribuie apoi la formarea eşantionului prin prelevări aleatoare, potrivit tehnicilor pe care le expunem în continuare

5.8.1 Sondaj tipic simplu

Se consideră o populaţie oarecare A, împărţită în s subpopulaţii parţiale, după un anumit criteriu de conţinut, ales judicios, cu respectarea principiilor unei grupări ştiinţifice:

$$A_1, \ A_2, \ ..., \ A_s.$$

Vom numi aceste subpopulaţii *grupe* sau *straturi*.

Fiecăreia din cele s grupe îi corespund următoarele valori ale caracteristicii

$$A_1 \ (a_{11}, \ a_{21}, \ ..., \ a_{N_1 1})$$
$$A_2 \ (a_{12}, \ a_{22}, \ ..., \ a_{N_2 2})$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$A_s \ (a_{1s}, \ a_{2s}, \ ..., \ a_{N_s s}).$$

După cum se vede, stratul A_1 are N_1 unităţi, stratul A_2 are N_2 unităţi etc. Numărul total al unităţilor populaţiei A este:

$$N_1 + N_2 +, \ ..., N_s = N.$$

Scanned with OKEN Scanner

Din fiecare dintre aceste straturi facem, respectiv, cîte

$$n_1, \ n_2, \ ..., \ n_s$$

extracţii la întîmplare, folosind în acest scop, de pildă, un tabel al numerelor aleatoare, fără ca unitatea extrasă să fie înapoiată în populaţia originară.

Vom obţine, evident, un eşantion format din n unităţi, însumînd toate extracţiile efectuate:

$$n_1 + n_2 +, ..., + n_s = n.$$

Prin urmare, extrăgînd din fiecare strat cîte un eşantion parţial n_j ($j = 1, 2, ..., s$), obţinem s asemenea eşantioane ale căror unităţi au caracteristici cu valorile respective:

$$\begin{array}{l} x_{11}, \ x_{21}, \ ... \ x_{n1} \\ x_{12}, \ x_{22}, \ ... \ x_{n2} \\ \vdots \ \ldots\ldots\ldots\ldots \\ x_{1s}, \ x_{2s}, \ ... \ x_{ns} \end{array}$$

unde x_{ij} ($i = 1, 2, ..., n_j$; $j = 1, 2, ..., s$) sînt de considerat, evident, ca fiind variabile aleatoare.

Introducînd notaţiile:

$$m = \frac{1}{N} \sum_{j=1}^{s} \sum_{i=1}^{N_j} a_{ij}$$

$$m_j = \frac{1}{N_j} \sum_{i=1}^{N_j} a_{ij},$$

rezultă că este valabilă egalitatea:

$$m = \frac{1}{N} \sum_{j=1}^{s} N_j m_j. \tag{5.27}$$

Deci media caracteristicii în populaţia originară poate fi interpretată ca fiind media ponderată a mediilor de grupă ale caracteristicii, ponderile fiind egale cu:

$$\frac{N_j}{N}, \ (j = 1, \ 2, ..., s).$$

În mod analog, notînd:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{s} \sum_{i=1}^{nj} x_{ij}$$

$$\bar{x}_j = \frac{1}{n_j} \sum_{j=1}^{s} x_{ij}$$

deducem egalitatea:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{s} n_j \bar{x}_j, \tag{5.28}$$

Scanned with OKEN Scanner

Deci, media valorilor din eşantionul general (de mărimea $\sum_{j=1}^{s} n_j = n$) este egală cu media ponderată a mediilor de grupă $\bar{x}_{js}$ ponderile fiind egale, respectiv cu:

$$\frac{n_j}{n}, \quad (j = 1, 2, ..., s).$$

Estimarea mediei din populaţia originară. Pentru estimarea mediei a din populaţia originară pe baza unui sondaj tipic simplu vom considera variabila y definită astfel:

$$y = \frac{1}{N} \sum_{j=1}^{s} N_j \bar{x}_j. \tag{5.29}$$

După cum se vede, această variabilă reprezintă o medie ponderată a rezultatelor $\bar{x}_j$, obţinute în fiecare grupă. Cu ajutorul ei putem estima media a în populaţia originară, considerînd următoarea teoremă:

Teorema 5.1. În cazul unui sondaj tipic simplu sînt valabile următoarele egalităţi:

$$M(y) = m \tag{5.30}$$

$$D^2(y) = \sum_{j=1}^{s} \frac{N_j^2}{N^2} (1 - f_i) \frac{s_j^2}{n_j}, \tag{5.31}$$

unde

$$s_j^2 = \frac{1}{N_j - 1} \sum_{j=1}^{N_j} (a_{ij} - a_j)^2, \; f_j = \frac{n_j}{N_j}. \tag{5.32}$$

Din această teoremă rezultă condiţiile de aplicare a sondajului tipic simplu. De asemenea, teorema indică o estimaţie absolut corectă a mediei a din populaţia originară.

Într-adevăr, dacă eşantioanele de grupă n_j $(j = 1, 2, ..., s)$ sînt suficient de mari, atunci dispersia $D^2(y)$ este foarte mică şi prin urmare se poate afirma că, în aceste condiţii, variabila constituie un estimator absolut corect al mediei în populaţia originară. Această observaţie are o deosebită importanţă practică.

Aplicaţia 7. Într-o întreprindere industrială în care lucrează un număr de 1200 muncitori s-a hotărît să se studieze, pe bază de sondaje, mărimea medie a întreruperilor în timpul de lucru al muncitorilor, în decurs de o zi. Dat fiind că mărimea întreruperilor în muncă variază de la o specialitate la alta a muncitorilor, s-a recurs la folosirea sondajului tipic simplu.

Populaţia muncitorilor a fost divizată în patru grupe, pe specialităţi, şi apoi s-au ales, la întîmplare, cîte 50 de muncitori din fiecare grupă şi s-a cronometrat timpul de lucru şi cel de întrerupere în fiecare eşantion parţial.

Scanned with OKEN Scanner

. Datele obținute pe cele patru grupe se văd în tabelul 5.5.

Tabelul 5.5

Specialitatea muncitorilor	Numărul muncitorilor în fiecare grupă a populației originare	Numărul muncitorilor cuprinși în sondaj	Proporția eșantionului de grupă	Numărul mediu de minute de întrerupere	Abaterea medie pătratică de grupă
s_j $(j = 1, 2, 3, 4)$	N_j	n_j	$\frac{n_j}{N_j} = f_j$	$\bar{x}_j$	s_j
0	1	2	3	4	5
Strungari	500	50	0,10	30	8
Frezori	200	50	0,25	40	10
Ajustori	150	50	0,33	20	7
Rectificatori	350	50	0,14	30	9
Σ	$1200 = N$	$200 = n$	0,16	$30,41 = \bar{x}$	

Vom avea:

$$y = \bar{x} = M(\bar{x}_j) = \frac{1}{N}\sum_{j=1}^{s} N_j \bar{x}_j = \frac{15\,000 + 8\,000 + 3\,000 + 10\,500}{1\,200} = 30,41 \text{ minute.}$$

Pentru a estima media în populația originară vom calcula eroarea standard a mediei de sondaj; ținînd seama de formula (5.31), indicatorii de calcul se văd în tabelul 5.6.

Tabelul 5.6

Specialitatea muncitorilor	$\frac{N_j}{N}$	$\left(\frac{N_j}{N}\right)^2$	$1 - f_i$	s_j^2	$\frac{s_j^2}{n_j}$	$\frac{N_j^2}{N^2}(1-f_j)\frac{s_j^2}{n_j}$
Strungari	0,41	0,168	0,90	64	1,28	0,193536
Frezori	0,17	0,028	0,75	100	2,00	0,042000
Ajustori	0,13	0,016	0,67	49	0,98	0,010505
Rectificatori	0,29	0,084	0,86	81	1,62	0,117028
	1,00	—	—	—	—	0,363069

Prin urmare, avem:

$$s'_{\bar{x}} = \sqrt{\Sigma \frac{N_i^2}{N^2}(1 - f_j)\frac{s_i^2}{n_j}} = \sqrt{0,363069} \simeq 0,60.$$

Pentru estimarea mediei în populația originară a, vom considera eroarea limită de sondaj cu o siguranță probabilistică de 95% căreia îi corespunde, după cum știm, factorul $d = 1,96$.

Scanned with OKEN Scanner

Deci:

$d \cdot s'_x = 1{,}96 \cdot 0{,}60 = 1{,}176 \approx 0{,}18$ minute, iar intervalul de încredere pentru localizarea mediei în populaţia originară va fi:

$$30{,}41 - 0{,}18 < m < 30{,}41 + 0{,}18,$$

ceea ce înseamnă că mărimea medie a întreruperii în timpul de lucru zilnic al unui muncitor, pe întreaga populaţie de 1200 muncitori, nu va fi mai mică de 29,43 minute şi nici mai mare de 31,59 minute şi această afirmaţie o facem cu o probabilitate de 95%.

5.8.2 Sondaj tipic proporţional

Specificul acestui fel de sondaj constă în faptul că din fiecare grupă tipică a populaţiei originare se extrage, la întîmplare, un anumit număr de unităţi astfel încît raportul dintre numărul lor şi volumul grupei din care s-au extras să fie egal cu raportul dintre volumul eşantionului global şi cel al populaţiei originare.

Prin definiţie deci, se numeşte sondaj tipic *proporţional* sau *reprezentativ*, un sondaj stratificat pentru care sînt îndeplinite condiţiile:

$$f_j = f, \quad (j = 1, \ 2, \ ..., \ s),$$

de unde

$$f = \frac{n_1}{N_1} = \frac{n_2}{N_2} = ... = \frac{n_s}{N_s} = \frac{\sum\limits_{j=1}^{s} n_j}{\sum\limits_{j=1}^{s} N_j} = \frac{n}{N},$$

din care deducem:

$$n_j = \frac{n}{N} N_j, \ (j = 1, \ 2, \ ..., \ s).$$

Dispersia funcţiei de estimaţie y, în cazul sondajului tipic proporţional, este:

$$D^2(y) = \frac{1 - f}{n} \sum_{j=1}^{s} \frac{N_j}{N} s^2, \tag{5.33}$$

ceea ce se poate uşor constata introducînd valoarea de mai sus a lui n_j în expresia dispersiei (5.31) şi (5.32).

Aplicaţia 8. Să considerăm din nou exemplul populaţiei generale de 1200 muncitori pentru care s-a calculat, pe bază de sondaj tipic, simplu, mărimea medie a întreruperii în timpul zilnic de lucru al unui muncitor. De data aceasta însă vom aplica metoda sondajului tipic proporţional, notînd volumul general al eşantionului $n = 200$ sau 16,66% (aproximativ 17%). Din fiecare grupă vor fi extrase, la întîmplare, eşantioane parţiale reprezentînd 16,66% din volumul grupei respective.

Scanned with OKEN Scanner

Tabelul 5.7

Specialitatea muncitorilor	$n_j = \frac{N_j}{N} s_j^2$	$\frac{N_j}{N}$	s_j^2	$\frac{N_j}{N} \cdot s_j^2$
0	1	2	3	4
Strungari	83	0,41	64	26,24
Frezori	34	0,17	100	17,00
Ajustori	25	0,13	49	6,37
Rectificatori	58	0,29	81	23,49
Σ	$200 = n$	1,00	—	73,10

În tabelul 5.7 sînt prezentate eşantioanele de grupă şi indicatorii necesari pentru calculul erorii standard.

După cum se ştie, eroarea standard a mediei de sondaj se obţine extrăgînd rădăcina pătrată din expresia (5.33)

$$s'_x = \sqrt{\frac{1 - f}{n} \sum_{j=1}^{s} \frac{N_j}{N} s_j^2}; \qquad (5.34)$$

Introducînd datele din tabelul 5.6. în această expresie, obţinem următorul rezultat:

$$s'_x = \sqrt{\frac{1 - \frac{200}{1200}}{200} \cdot 73{,}10} = \sqrt{0{,}304577} = 0{,}552 \text{ minute.}$$

De data aceasta, după cum se vede, am obţinut o eroare standard a mediei de sondaj mai mică decît atunci cînd am aplicat metoda sondajului tipic simplu (0,552 faţă de 0,60 minute).

Cu o probabilitate de 0,95 vom avea următoarea inegalitate:

$$\bar{x} - 1{,}96 \cdot 0{,}552 \leqslant M(X) \leqslant \bar{x} + 1{,}96 \cdot 0{,}552.$$

Este evident că, dat fiind că acest rezultat s-a obţinut cu acelaşi nivel de semnificaţie şi cu acelaşi volum general al eşantionului ($n = 200$) ca şi în primul caz, putem conchide că în cazul sondajului tipic proporţional lucrăm şi cu mai multă precizie.

În legătură cu eficienţa sondajului tipic proporţional se impun două observaţii:

1°. Din punct de vedere teoretic, nu putem trage o concluzie generală în sensul că sondajul tipic simplu este mai puţin eficient decît sondajul tipic proporţional. Aceasta înseamnă că din punct de vedere teoretic se pot construi planuri de sondaj tipice, simple, care să fie, totuşi, mai eficiente decît sondajele tipice proporţionale.

Se poate arăta că pentru aceasta este suficient să aibă loc inegalitatea:

$$n_j \leqslant N_j - \frac{N - n}{n}, \quad (j = 1, 2, 3, \ldots, s) \qquad (5.35)$$

şi toate valorile dispersiilor s_j^2 să fie foarte mici.

Scanned with OKEN Scanner

Din punct de vedere practic însă, remarcăm că inegalitatea (5.35) nu este niciodată satisfăcută, întrucît ar trebui ca fiecare N_j să fie mai mic decît 100 unităţi ($N < 100$), ceea ce nu se întîmplă în realitate, dacă ţinem seama că metoda sondajului nu se aplică decît atunci cînd numărul unităţilor populaţiei originare este foarte mare.

2°. *Sondajul tipic proporţional este întotdeauna mai eficient decît sondajul tipic simplu întrucît dă naştere la o dispersie mai mică.*

Dacă considerăm două eşantioane în felul următor: un eşantion aleator de n unităţi extras din populaţia A, iar al doilea, de aceeaşi mărime, format însă pe baza unor eşantioane de grupă, aplicînd sondajul tipic proporţional la aceeaşi populaţie A, împărţită în s grupe.

Se poate demonstra că pentru aceste două eşantioane avem următoarea inegalitate a dispersiilor:

$$\frac{1}{n}(1-f)s^2 \geqslant \frac{1}{n}(1-f)\sum_{j=1}^{s}\frac{N_j}{N}s_j^2 \tag{5.36}$$

exprimînd tocmai faptul că sondajul tipic proporţional este mai eficient decît sondajul tipic simplu.

5.8.3 Sondaj tipic optim

Problema fundamentală care se pune în cadrul acestui tip de sondaj este aceea de a fi determinate eşantioane de grupă n_j, ($j = 1, 2, ..., s$) astfel încît eficienţa sondajului să fie *maximă*.

În teoria sondajului se demonstrează că o asemenea determinare este posibilă.

Aceasta revine la a face minimă expresia (5.31) condiţionată de egalitatea:

$$n_1 + n_2 + , ..., + n_s = n.$$

Se ajunge astfel la concluzia că trebuie să avem eşantioanele de grupă n_j, determinate pe baza următoarei expresii:

$$n_j = \frac{nN_js_j}{\sum_{j=1}^{s}N_js_j}, \quad (j = 1,\ 2,\ 3, ..., s) \tag{5.37}$$

asigurîndu-se astfel eficienţa maximă a sondajului.

Un asemenea sondaj poartă numele de sondaj tipic *optim*.

Eficienţa sondajului tipic optim. Pentru a cunoaşte eficienţa acestui sondaj este necesar să calculăm dispersia funcţiei de estimaţie (5.29) corespunzătoare acestui tip de sondaj. Ea se obţine înlocuind valoarea lui n_j din (5.37) în expresia $D^2(y)$ dată de (5.31).

$$D^2(y) = \sum_{j=1}^{s}\frac{N_j^2}{N^2}\left(1-\frac{n_j}{N_j}\right)\frac{s_j^2}{n_j} = \sum_{j=1}^{s}\frac{N_j^2s_j^2}{N^2}\left(\frac{1}{n_j}-\frac{1}{N_j}\right) =$$

$$= \sum_{j=1}^{s}\frac{N_j^2\ s_j^2}{N^2}\left(\frac{\sum_{j=1}^{s}N_js_j}{ns_j}-1\right) = \frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{s}N_js_j\right)^2}{n} - \sum_{j=1}^{s}N_js_j^2\right]. \tag{5.38}$$

Scanned with OKEN Scanner

Aplicația 9. Să considerăm din nou exemplul folosit în cazul sondajelor tipice simple și proporționale, de data aceasta efectuînd însă un sondaj tipic optim.

În acest scop vom alege numerele n_j potrivit formulei (5.37). Elementele necesare le găsim în tabelul 5.8.

Tabelul 5.8

Specialitatea muncitorilor	$N_j s_j$	$\frac{N_j s_j}{\sum_{j=1}^{s} N_j s_j}$	$n_j \frac{m N_j s_j}{\sum_{j=1}^{s} N_j s_j}$
0	1	2	3
Strungari	4 000	0,39	78
Frezori	2 000	0,19	38
Ajustori	1 050	0,11	22
Rectificatori	3 150	0,31	62
	10 200	1,00	200 = n

În cazul sondajului tipic optim, eroarea standard a mediei de sondaj o obținem, evident, extrăgînd rădăcina pătrată din expresia (5.38)

$$s'_{\bar{x}} = \sqrt{D^2(y)} = \sqrt{D^2(\bar{x})} = \sqrt{\frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{s} N_j s_j\right)^2}{n} - \sum_{j=1}^{s} N_j s_j\right]}. \qquad (5.39)$$

Calculînd datele de sub radical, vom obține, în final, următorul rezultat:

$$s'_{\bar{x}} = \sqrt{\frac{1}{1\,200^2}\left[\frac{104\,040\,000}{200} - 87700\right]} = \sqrt{0{,}0301} \simeq 0{,}0301 \text{ minute}$$

și

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{s} n_j \bar{x}_j = \frac{(78 \cdot 40) + (38 \cdot 40) + (32 \cdot 20) + (62 \cdot 30)}{200} = \frac{6160}{200} = 30{,}8 \text{ m.}$$

Atragem atenția din nou că această medie este o variabilă aleatoare; ea diferă de la un sondaj la altul nedepășind — cu probabilitatea de 0,95, în cazul normalității repartiției — limitele:

$$\bar{x} \pm 1{,}96 \cdot s'_{\bar{x}}, \text{ adică } 30{,}8 \pm 1{,}96 \cdot 0{,}0301.$$

Prin urmare, mărimea medie a întreruperii în timpul de lucru zilnic al unui muncitor în populația originară nu va fi mai mică de 30,74 minute și nici mai mare de 30,86 minute.

Din exemplul considerat a rezultat în mod clar că *sondajul tipic optim este mai eficient decît sondajul tipic proporțional* și se poate demonstra, în general, că această afirmație are caracterul unei teoreme generale, întrucît dispersia mediei de sondaj, în cazul sondajului tipic optim este întotdeauna mai mică decît dispersia în cazul sondajului tipic proporțional.

Scanned with OKEN Scanner

5.9. SONDAJ ÎN FAZE

Prin sondaj în faze se înţelege acea cercetare pe bază de sondaj în care unităţile prelevate în eşantion constituie obiectul selectării în două sau mai multe faze.

În cele ce urmează, ne vom opri — spre ilustrare — asupra unui sondaj în două faze, vizînd o singură caracteristică, a populaţiei originare.

Să considerăm o populaţie originară oarecare A, împărţită în k grupe — straturi:

$$A_1, A_2, ..., A_k.$$

Într-o primă fază se aleg, la întîmplare, din cele k grupe, un număr de h grupe ($h < k$) pe care le vom numi, pentru mai multă claritate, unităţi *primare* de sondaj.

Mai departe, în faza următoare, din cele h grupe prelevate se fac extracţii aleatoare de unităţi elementare de sondaj — le vom numi unităţi-*secundare* —, procedînd în acelaşi mod ca şi în cazul sondajului tipic simplu. Dacă $h = k$ este clar că sîntem chiar în cazul sondajului tipic simplu, ceea ce pune în evidenţă faptul că acest din urmă sondaj este un caz particular al sondajului în două faze.

Este evident că aceleaşi probleme care se pun pentru sondajul tipic simplu se pun şi pentru sondajul în două faze. Sondajul în două faze se impune de preferinţă atunci cînd ni se pune problema ca într-un timp relativ scurt să obţinem informaţiile statistice de care avem nevoie. De asemenea, putem recurge la un sondaj în două faze dacă costul total al cheltuielilor pentru efectuarea sondajului într-o singură fază ar fi mult prea mare faţă de posibilităţile de care dispunem.

De pildă, în România există 2706 de comune rurale*. Ni se cere, pe baza unui sondaj, să estimăm numărul de familii avînd fiecare mai mult de cinci copii. Considerînd că fiecare comună rurală reprezintă o grupă de familii, putem estima numărul de familii pe care îl căutăm, recurgînd la un sondaj tipic, simplu. După cum se ştie, potrivit metodologiei acestui sondaj vom lua la întîmplare din fiecare comună cîte un eşantion de familii, reprezentînd fiecare, să zicem, 10% din numărul familiilor din fiecare comună şi apoi pe baza eşantionului global astfel format, vom stabili numărul absolut de familii care au mai mult de cinci copii. Este clar că metoda sondajului tipic, simplu, ne va cere un timp de lucru mai îndelungat, întrucît implică prelevarea unui număr mare de unităţi de observare.

De aceea, în exemplul considerat putem recurge la un sondaj în două faze Vom alege, deci, la întîmplare, într-o primă fază, un număr de 100 comune rurale şi apoi, în faza a doua, vom preleva, din fiecare din aceste 100 comune, eşantioane de familii, reprezentînd cîte 10% din numărul familiilor din comunele respective. Din punctul de vedere al economicităţii cercetării întreprinse, avantajul sondajului în două faze este incontestabil.

Vom sublinia că, deseori în practică, s-au folosit sondaje în două faze şi în alt scop, specifică fiind următoarea situaţie: în prima fază sînt studiate *toate* unităţile populaţiei de sondaj vizîndu-se, ca prim obiectiv, anumite caracteristici bine determinate, în faza a doua sînt cercetate numai *o parte*

* Vezi *Anuarul statistic al R. S. România 1975*, p. 5.

din unitățile care au fost prelevate în prima fază și aceasta cu scopul de a fi studiate și unele caracteristici suplimentare.

Așa, de pildă, dacă în prima fază a unui sondaj au fost studiate 5 000 de familii, înregistrîndu-se cu acest prilej o singură caracteristică, și anume nivelul venitului fiecărei familii, în a doua fază, se poate alege, la întîmplare, un număr de 10% din familii din cele 5 000 de familii cercetate în prima fază, de data aceasta însă înregistrîndu-se și o serie de alte caracteristici suplimentare cum ar fi, de pildă cele referitoare la sursa venitului, destinația lui pe diferite categorii de cheltuieli (consum, îmbrăcăminte etc.)

Rezolvarea matematică a problemei este aceeași, într-un caz și în altul.

5.9.1 Sondaj în două faze

La fel ca și la sondajul într-o singură fază și cu o singură caracteristică vom nota cu

$$x_{ij},\ (i = 1,\ 2, ..., n_j)$$

valorile ieșite pentru unitățile secundare din grupa $A_j (j = 1,\ 2, ..., k)$ și cu Y_j, expresia:

$$Y_j = \frac{kN_j}{N} \cdot \frac{\sum_{i=1}^{n_j} x_{ij}}{n}, \tag{5.40}$$

Formula (5.40) trebuie să capete o justificare.
Expresia:

$$\frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

reprezintă media corespunzătoare pentru grupa j. Ar fi inexact s-o considerăm ca medie pentru întreaga populație, deoarece trebuie să ținem seama de raportul unităților din această grupă și celelalte grupe.

Dacă toate grupele ar avea N_j unități am avea un număr total de kN_j unități. Dacă $kN_j > N$, înseamnă că grupa N_j este o grupă mare; dacă $kN_j < N$, înseamnă că avem o grupă mică. Raportul $\frac{kN_j}{N}$ reprezintă un indicator al volumului unităților din A_j față de celelalte grupe. Media obținută mai sus este natural deci s-o luăm cu ponderea $\frac{kN_j}{N}$ și obținem pentru stratul A_j valoarea (5.40)

Să observăm că dacă

$$N_j = \frac{N_j}{k},$$

atunci ponderea $\frac{kN_j}{N}$ ia valoarea 1. Expresia Y_j variază de la grupă la grupă și este o variabilă aleatoare.

Scanned with OKEN Scanner

Fie indicatorul

$$Y = \frac{Y_{j_1} + Y_{j_2} + \ldots + Y_{j_h}}{h} \tag{5.41}$$

indicii $j_1, j_2, \ldots, j_h$ simbolizînd grupele care ne-au ieşit la întîmplare — unităţile primare — după executarea primei faze a sondajului.

Folosind aceleaşi notaţii ca în expunerea precedentă, putem enunţa următoarea teoremă:

Teorema 5.2. Într-un sondaj în două faze avem:

$$M(Y) = m. \tag{5.42}$$

$$D^2(Y) = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\alpha_j - \alpha)^2}{k-1} + \frac{k}{hN^2} \sum_{j=1}^{k} N_j^2 (1 - f_j) \frac{s_j}{n_j}; \tag{5.43}$$

$$m = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{N_j} a_{ij} \tag{5.44}$$

$$\alpha_j = \sum_{i=1}^{N_j} \alpha_{ij} \tag{5.45}$$

$$\alpha = \sum_{j=1}^{k} \alpha_j. \tag{5.46}$$

Vom aminti că simbolurile f_j şi s_j^2 au aceleaşi semnificaţii ca în teorema 5.1. din expunerea precedentă privind sondajul tipic, adică:

$$f_j = \frac{n_j}{N_j}; \quad s_j^2 = \frac{1}{N_j - 1} \sum_{j=1}^{N_j} (a_{ij} - a_j)^2.$$

Simbolul α_j se referă la suma valorilor pe care le ia caracteristica cercetată la unităţile secundare de sondaj.

Demonstraţia acestei teoreme se obţine uşor dacă ţinem seama de rezultatele prezentate în contextul expunerilor privind teoria sondajului într-o singură fază.

Teorema de mai sus ne oferă suportul ştiinţific pentru aplicarea în practică a sondajului în două faze.

Pentru ca rezultatele obţinute să fie exacte este necesar — în virtutea inegalităţii lui Cebîşev — ca dispersia lui Y dată de relaţia (5.43) să fie foarte mică. Vom examina cazul în care acest lucru se întîmplă.

După cum se vede din (5.43), dispersia variabilei aleatoare Y este suma a doi termeni

$$D_1 = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\alpha_j - \alpha)^2}{k-1}$$

şi

$$D_2 = \frac{k}{hN^2} \cdot \sum_{j=1}^{k} N_j^2 (1 - f_j) \frac{s_j^2}{n_j}.$$

Scanned with OKEN Scanner

Primul termen, D_1, reprezintă dispersia corespunzătoare primei faze a sondajului şi se referă la suma valorilor pe care le ia caracteristica cercetată la unităţile secundare de sondaj:

$$\alpha_j = \sum_{j=1}^{N_j} \alpha_{ij}.$$

Termenul D_1 îl putem considera deci ca fiind dispersia între grupe; el se anulează dacă $h = k$, adică atunci cînd sondajul încetează de a mai fi în două faze şi se reduce la un sondaj grupat, simplu. De asemenea, acest termen se anulează şi atunci cînd avem egalităţile:

$$\alpha_1 = \alpha_2 = \ldots = \alpha_k;$$

cu alte cuvinte, atunci cînd suma valorilor pe care le ia caracteristica cercetată la unităţile secundare este aceeaşi pentru fiecare grupă.

Al doilea termen, D_2, reprezintă dispersia din interiorul grupelor care corespunde fazei a doua a sondajului. El se anulează cînd avem $n_j = N_j$ $(j = 1, 2, \ldots, k)$, adică atunci cînd din fiecare grupă k sînt scoase toate unităţile secundare.

Rezultă că dacă numerele n_j sînt suficient de mari, termenul $D_2 \to 0$, în timp ce termenul D_1 rămîne constant, întrucît nu depinde de n_j.

Referitor la cele expuse putem face următoarea observaţie:

Variabila aleatoare

$$Y = \frac{Y_{j1} + Y_{j2} + \ldots + Y_{jh}}{h}$$

nu este o funcţie de estimaţie pentru media $M(Y)$ a caracteristicii studiate în populaţia originară, oricît de mare ar fi volumul eşantionului

$$n = \sum_{j=1}^{k} n_j.$$

Pentru ca din punct de vedere practic să ajungem la rezultate acceptabile şi în cazul unui sondaj în două faze, este necesar ca dispersia $D^2(Y)$ să fie foarte apropiată de zero, pentru numerele $n_j (j = 1, 2, \ldots, k)$ suficient de mari. Acest rezultat nu se poate obţine însă decît numai dacă D_1, la rîndul său, este foarte apropiat de zero.

Or, pentru ca D_1 să fie egal cu zero trebuie să avem unul din următoarele cazuri:

$$h = k \text{ sau } \alpha_j = \alpha.$$

Primul caz nu interesează, deoarece avantajul sondajului în două faze constă tocmai în faptul că avem posibilitatea de a alege, la întîmplare, un număr mai mic de unităţi primare decît cel existent.

Prezintă deci interes deosebit rezolvarea cazului al doilea, cînd $\alpha_j = \alpha (j = 1, 2, \ldots, k)$. Este însă clar că, spre deosebire de sondajul tipic simplu, sondajul în două faze nu este aplicabil decît după o prealabilă prelucrare şi formare judicioasă a grupelor.

Scanned with OKEN Scanner

Pentru a pune în evidență erorile ce se pot ivi în aplicarea, fără un control prealabil, a sondajului în două faze, vom considera, într-o formă schematică, următoarea aplicație:

Aplicația 10. Să presupunem că avem $k = 4$ și că, în planul nostru de sondaj, am acceptat $h = 2$, adică populația originară are patru grupe, din care ne-am propus să extragem, la întîmplare, două.

Mai admitem că populația originară se caracterizează prin următoarele date privind volumul ei total și pe cele patru grupe ca și suma valorilor caracteristicii:

$N_1 = 10\,000$	$\alpha_1 = 100\,000$
$N_2 = 12\,000$	$\alpha_2 = 25\,000$
$N_3 = 10\,000$	$\alpha_3 = 200\,000$
$N_4 = 18\,000$	$\alpha_4 = 75\,000$
$N = 50\,000$	$\alpha = 400\,000$

După cum se vede, valorile α_j diferă sensibil între ele.

Media caracteristicii în populația originară este, evident, egală cu:

$$m = \frac{\alpha}{N} = \frac{400\,000}{50\,000} = 8.$$

De asemenea, mai admitem că, după ce am ales cele două unități primare, lucrăm în cele mai exacte condiții; adică, din fiecare grupă, scoatem toate unitățile secundare. În aceste condiții, aplicînd formulele stabilite, rezultă pentru media m, în populația originară, următoarele estimații:

Faza a doua a sondajului s-a făcut din grupele primare:	Estimația mediei
A_1, A_2	5
A_1, A_3	12
A_1, A_4	7
A_2, A_3	9
A_2, A_4	4
A_3, A_4	11

Estimația medie se face după formula (5.40)

Pentru grupa A_1, avem ținînd seama că

$$k = 4,\ N_1 = 10\,000,\ \alpha_1 = 100\,000,\ n_1 = 10\,000,\ \sum x_{i1} = 100\,000$$

$$Y_1 = 4\,\frac{10\,000}{50\,000} \cdot \frac{100\,000}{10\,000} = 8.$$

Pentru grupa A_2:

$$k = 4,\ N_2 = 12\,000,\ \alpha_2 = 25\,000,\ n_2 = 12\,000,\ \sum x_{i2} = 25\,000$$

$$Y_2 = 4\,\frac{12\,000}{50\,000} \cdot \frac{25\,000}{12\,000} = 2.$$

Estimația o obținem aplicînd formula (5.41)

$$\frac{Y_1 + Y_2}{2} = \frac{8 + 2}{2} = 5.$$

La fel se obțin și celelalte estimații.

Scanned with OKEN Scanner

Valoarea exactă a mediei fiind 8, se vede din exemplul considerat că aplicarea sondajului în două faze, chiar și în cele mai bune condiții ($n_j = N_j$), ne conduce la rezultate inacceptibile din punct de vedere practic.

Folosind același exemplu, vom considera o altă ipoteză cu privire la valorile α_j.

Astfel, vom presupune că avem:

$N_1 = 10\,000$	$\alpha_1 = 100\,000$
$N_2 = 12\,000$	$\alpha_2 = 90\,.000$
$N_3 = 10\,000$	$\alpha_3 = 110\,000$
$N_4 = 18\,000$	$\alpha_4 = 100\,000$
$N = 50\,000$	$\alpha = 400\,000$

Se vede că acum valorile α_j nu diferă mult între ele și de aceea, teoretic, ne putem aștepta la estimații mult mai apropiate de realitate.

Media caracteristicii în populația originară este aceeași:

$$m = \frac{\alpha}{N} = 8.$$

Lucrînd în aceleași condiții, adică alegînd la întîmplare două grupe din cele patru și extrăgînd din fiecare grupă toate unitățile secundare, teoretic putem obține oricare din estimațiile de mai jos ale mediei m, în populația originară:

Faza a doua a sondajului s-a făcut în grupele primare:	Estimația mediei
A_1, A_2	7,6
A_1, A_3	8,8
A_1, A_4	8,0
A_2, A_3	8,0
A_2, A_4	7,6
A_3, A_4	8,8

Aceste rezultate sînt mult mai apropiate de valoarea exactă a mediei aritmetice în populația originară. Este evident că dacă am lua toate valorile $\alpha_j = 100\,000$ ($j = 1, 2, 3, 4$) estimația mediei m ar fi chiar egală cu 8.

Această observație ne arată că pentru aplicarea unui sondaj în două faze este necesară o prealabilă stratificare judicioasă a populației originare astfel încît media valorilor pe care le ia caracteristica cercetată, în cadrul fiecăreia din grupele formate, să fie aproximativ aceeași. Valoarea practică deosebită a acestei indicații metodologice nu comportă nici un fel de discuție.

5.9.2 Eroarea standard a mediei în cazul sondajului în două faze

Prin analogie cu metodologia folosită la definirea erorilor standard ale diferitelor valori tipice de sondaj în cazul sondajelor într-o singură fază, vom considera abaterea medie pătratică a mediei de sondaj în două faze drept eroarea ei standard. Concret, ea se obține extrăgînd rădăcina pătrată din expresia (5.43)

$$s_{\bar{x}}^2 = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\alpha_j - \alpha)^2}{k-1} + \frac{k}{hN^2} \cdot \sum_{j=1}^{k} N_j^2 (1 - f_j) \frac{s_j^2}{n_j}. \qquad (5.47)$$

Cu o probabilitate mai mare de 0,89 se poate afirma deci că media $M(Y)$ a populației originare se va găsi în intervalul

$$\bar{x} \pm 3 \cdot s'_{\bar{x}}.$$

O primă concluzie ce se desprinde de cele expuse pînă acum se referă la precizia sondajului în două faze, vizînd o singură caracteristică. Anume, se poate conchide că eroarea sondajului în două faze rezultă din suma erorilor celor două faze ale acestuia și că este mai mare decît oricare dintre acestea. Așadar, în comparație cu celelalte tipuri de sondaj, sondajul în două faze este mai puțin precis dat fiind că în faza a doua sporește eroarea totală.

Este limpede că în cazul mai multor faze, eroarea totală a sondajului crește considerabil. De aceea, numărul exagerat al fazelor sondajului trebuie evitat, știut fiind că fiecare fază nouă mărește eroarea totală.

O altă concluzie se referă la avantajele sondajului în faze:

a) sondajul în faze are o serie de avantaje derivînd din economicitatea ce-i este caracteristică și înlesnirile cu caracter organizatoric pe care le poate implica, mai ales atunci cînd obiectul cercetării are, de pildă, o mare extindere teritorială;

b) mediile se pot obține pe fiecare grupă inclusă în sondaj, la fel ca și în cazul sondajelor tipice;

c) în cazul sondajului în faze, pot fi ușor eliminate atît erorile de înregistrare, cît și erorile grosolane de reprezentativitate, asigurîndu-se astfel obținerea unor informații și apoi concluzii adeseori mai juste decît în cazul în care s-ar aplica alte tehnici de sondaj.

5.10. SONDAJ SIMPLU PRIVIND O CARACTERISTICĂ-ATRIBUT

În acest paragraf ne vom opri asupra problemelor teoretice și metodologice pe care le ridică sondajul care vizează un alt tip de caracteristici decît cele care variază cantitativ, așa-numitele *caracteristici-atribut**.

5.10.1 Considerații introductive

Prin caracteristici-atribut vom înțelege anumite *proprietăți* ale unităților populației cercetate, legate de natura lor specifică și care se exprimă, de obicei, printr-o indicație *calitativă*. Luînd în studiu o asemenea caracteristică ne interesează, în primul rînd, dacă un element oarecare al populației, inclus în eșantion, o posedă sau nu o posedă. Adeseori o caracteristică-atribut capătă forma unei variabile alternative, a cărei repartiție teoretică este în general binominală sau hipergeometrică (schema bilei revenite sau nerevenite).

În esență, procesul sondajului simplu asupra unei caracteristici-atribut urmează schema extragerii succesive a mai multor eșantioane dintr-o populație care conține elemente A și elemente non A.

Numărul absolut de elemente A în fiecare eșantion sau proporția de elemente A vor fi deci datele pe care le furnizează eșantioanele alese, cercetă-

* Principalele idei din această secțiune sînt expuse după [134].

Scanned with OKEN Scanner

torul urmînd a le extinde, în mod adecvat, asupra populației originare din care au fost extrase.

Adoptînd terminologia folosită în teoria probabilităților, vom considera că extragerea unui element în procesul sondajului este un *eveniment*. Apariția unui element A o vom numi eveniment *favorabil*, iar neapariția lui — adică apariția unui element non A, — ca eveniment *nefavorabil*.

De pildă, în procesul cercetării prin sondaj a calității produselor industriale ne poate interesa numărul sau proporția produselor rebut. În acest scop putem extrage un eșantion de 100 produse și să cercetăm pe fiecare în parte pentru a vedea dacă corespunde sau nu corespunde prescripțiilor tehnice de calitate: drept rezultat putem constata, de exemplu, că un număr de 5 produse nu îndeplinesc aceste condiții.

Eșantionul ne sugerează astfel ideea că în populația produselor cercetate, probabil aproximativ 5% sînt produse rebut.

Potrivit terminologiei convențional adoptate vom spune că eșantionul nostru ne-a oferit — sub aspectul caracteristicii-atribut luată în studiu — o proporție de 5% cazuri favorabile și 95% cazuri nefavorabile.

5.10.2 Media și eroarea standard a caracteristicii-atribut în cazul sondajului simplu

Să presupunem că am prelevat, la întîmplare, k eșantioane, fiecare conținînd n evenimente. Probabilitatea unui eveniment favorabil este, de fiecare dată, p, iar a unui eveniment nefavorabil $q(q = 1 - p)$.

După cum știm, frecvențele teoretice ale eșantioanelor cu 0, 1, 2, ... evenimente favorabile sînt termenii binomului $k(p + q)^n$, adică:

$$k\left(p^n + np^{n-1}q + \frac{n(n-1)}{2!}p^{n-2}q^2 + \ldots + npq^{n-1} + q^n\right).$$

Se vede că este vorba de o repartiție binomială.

Vom avea deci pentru $k = 1$ următoarele valori tipice ale repartiției *numărului absolut* de cazuri favorabile:

Media

$$M = np. \tag{5.48}$$

Eroarea standard a mediei numărului absolut de cazuri favorabile:

$$s'_{np} = \sqrt{npq}. \tag{5.49}$$

Dacă pe baza datelor de sondaj vom considera nu numărul cazurilor favorabile, ci *proporția* lor, împărțind toate rezultatele înregistrate prin n — volumul eșantionului —, avem de a face cu o altă variabilă aleatoare binominală caracterizată prin următoarele valori tipice:

Media

$$M = p. \tag{5.50}$$

Eroarea standard a proporției medii, a cazurilor favorabile

$$s'_p = \sqrt{\frac{pq}{n}}. \tag{5.51}$$

Scanned with OKEN Scanner

În cercetarea, pe bază de sondaj, a caracteristicilor-atribut, relațiile (5.49) și (5.51) sînt fundamentale. Este de subliniat însă că în rezolvarea numeroaselor probleme practice, probabilitatea cazului favorabil de apariție a unui eveniment nu este cunoscută apriori astfel încît sîntem nevoiți să facem anumite considerații speciale înainte de a putea aplica aceste relații.

Pentru a preciza ideile vom considera următoarea aplicație:

Aplicația 11. Să considerăm o populație de oi dintr-un județ A din care se alege, la întîmplare, un număr de 900 și se constată că 180 din ele sînt oi merinos, adică 20%. În mod firesc ne punem întrebarea ce putem spune despre proporția de oi merinos în întreaga populație de oi a județului considerat.

În acest caz, nu cunoaștem proporția teoretică, p, a oilor merinos. Știm însă că eroarea standard s'_{np} este egală cu $\sqrt{900\,p\cdot q}$. Se poate afirma că oricare ar fi p și q, produsul pq nu depășește 1/4 și deci eroarea standard nu depășește $1/2\sqrt{900}$, adică 15 cazuri.

Prin urmare, oricare ar fi p, un sondaj simplu ar trebui să ne dea un număr de cazuri favorabile care să se abată de la frecvența teoretică cu mai puțin de $3\cdot 15 = 45$ cazuri. Acest număr reprezintă 5% din eșantion și astfel am putea trage concluzia că proporția oilor merinos în întreaga populație — exprimată în procente — este de $20\% \pm 5\%$, adică se află în intervalul de 15% și 25%.

Desigur se poate obține o estimație ceva mai bună a acestei proporții. Am văzut că eroarea standard este mică în comparație cu valoarea medie; rezultă că ea este mai mică în comparație cu valoarea observată.

Prin urmare, dacă calculăm eroarea standard pe baza valorilor observate în eșantion, le vom nota cu w și $1-w$ — în locul valorilor teoretice necunoscute p și q ($q = 1-p$), nu vom comite o eroare prea mare.

În exemplul de mai sus avem $w = 0{,}20$ și $1-w = 0{,}80$ și pe baza acestor date putem calcula eroarea standard:

$$s'_{np} = \sqrt{900\cdot 0{,}20\cdot 0{,}80} = \sqrt{144} = 12.$$

Considerînd această mărime putem spune că avem aproximativ $3\cdot s'_{np} = 36$ cazuri favorabile, reprezentînd 4% din eșantion.

Limitele proporției oilor merinos în întreaga populație sînt acum de $20\% \pm 4\%$, adică 16% și 24%, ceva mai înguste decît cele obținute în calculul anterior.

Aplicînd direct relația (5.51) care definește eroarea standard a proporției cazurilor favorabile (proporția oilor merinos) în întreaga populație ajungem la următorul rezultat:

$$s'_p = \sqrt{\frac{np}{n}} = \sqrt{\frac{0{,}2(1-0{,}2)}{900}} = \sqrt{\frac{0{,}16}{900}} = \frac{0{,}4}{30} \simeq 1{,}3\%.$$

El ne arată că proporția p — exprimată în procente — a oilor merinos în întreaga populație va fi cuprinsă considerînd $d = 3$ în următorul interval:

$$p \simeq w \pm 3\cdot 1{,}3$$

$$p \simeq 20\% \pm 3{,}6\%.$$

Scanned with OKEN Scanner

După cum se vede, limitele intervalului de încredere de 16,4% și 23,6% în care se va găsi proporția oilor merinos în întreaga populație, sînt determinate și mai precis, însă ele nu diferă prea mult de cele stabilite anterior.

În legătură cu această primă aplicație pe care am prezentat-o, sînt de făcut două observații, și anume:

Observația 1. Repartiția binomială a cărei importanță practică deosebită a fost pusă în evidență tocmai prin cercetarea — pe bază de sondaj — a caracteristicilor-atribut aparține tipului unimodal și aproximează foarte bine repartiția normală pentru valori mari ale lui n.

În marea majoritate a eșantioanelor vom constata deci valori absolute ale numărului de cazuri favorabile situate într-un interval delimitat de $3 \cdot s_{np}$ de fiecare parte a mediei $n \cdot w$, adică în următorul interval:

$$n \cdot w - 3\sqrt{n \cdot w(1-w)} \quad \text{și} \quad n \cdot w + 3 \cdot \sqrt{n \cdot w(1-w)}.$$

Dacă repartiția este exact normală, 0,9973 de valori se găsesc în acest interval. Se poate spune, prin urmare, că dacă un eșantion particular oarecare indică un număr de cazuri favorabile în afara acestui interval, este puțin probabil că abaterea de la valoarea teoretică să fie o consecință a fluctuațiilor de sondaj. Dacă n este mare, probabilitatea unui astfel de eșantion nereprezentativ este de numai 0,003. Subliniem din nou că regula celor trei erori standard este justificată dacă n este mare.

Observația 2. În exemplul considerat, am admis — oarecum intuitiv — că proporția cazurilor favorabile în eșantion (w), estimează proporția cazurilor favorabile în întreaga populație și am obținut astfel limitele intervalului de încredere în care se găsește valoarea teoretică a acestei proporții. Fără a dezvolta aici argumentările teoretice de rigoare precizăm că, într-adevăr, valoarea lui w în eșantion poate fi luată drept estimație a lui p în întreaga populație; de asemenea, trebuie să punem în evidență premisa ce o admitem cu acest prilej; în general, premisa că estimația w poate fi folosită în calculul erorii standard este justificată dacă n este mare și dacă nici p, nici q nu sînt mici.

Dacă eroarea standard a proporției cazurilor favorabile $s'_p = \sqrt{\frac{npq}{n}}$ este mică, atunci și diferența $w-p$ trebuie să fie mică și deci nu se va comite o eroare gravă introducîndu-se în calcule estimația

$$\sqrt{\frac{w(1-w)}{n}}.$$

Aplicația 12. Se consideră un eșantion de 900 zile din activitatea unei întreprinderi industriale și se constată că în 100 zile s-au înregistrat defecțiuni în alimentarea cu energie electrică.

Ne punem întrebarea: care sînt limitele probabile ale procentului de zile marcate prin defecțiuni în alimentarea cu energie electrică a întreprinderii date?

Luînd $w = \frac{1}{9}$ ca estimație a proporției p de zile marcate prin defecțiuni, găsim eroarea standard a acestei proporții:

$$s'_p = \sqrt{\frac{pq}{n}} \simeq \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{1}{9} \cdot \frac{8}{9} \cdot \frac{1}{900}} = 0{,}0105 \text{ sau } 1{,}05\%,$$

iar limitele intervalului de încredere pentru $d = 3$ sînt: $11{,}11\% \pm 3{,}15$, adică de aproximativ 8% și 14,25%.

Scanned with OKEN Scanner

Am subliniat mai sus importanța deosebită pe care o au relațiile (5.49) și (5.51) care definesc, respectiv, eroarea standard a numărului absolut de cazuri favorabile și a proporției acestora în eșantion. În legătură cu aplicarea lor în practică vom reține atenția și asupra altor aspecte interesante ce pot intra în considerare.

Dacă una dintre proporțiile p și q este foarte mică, relația (5.49) poate fi pusă sub forma aproximativă, deosebit de utilă. Să presupunem că p este foarte mic, astfel încît îl putem neglija pe p^2 cînd îl comparăm cu p.

De aceea, putem scrie următoarea egalitate aproximativă:

$$pq = p - p^2 \simeq p.$$

În consecință, relația (5.49) poate fi aproximată astfel:

$$s'_{np} = \sqrt{np} = \sqrt{M(x)}. \tag{5.49'}$$

Cu alte cuvinte, dacă proporția cazurilor favorabile este mică, abaterea standard a numărului de cazuri favorabile este egală, aproximativ, cu rădăcina pătratică a mediei numărului de cazuri favorabile. Din această constatare rezultă că eroarea standard o putem găsi chiar dacă p este necunoscut, *cu condiția ca cel puțin să știm că este mic.*

De fapt, ne găsim în cazul în care repartiția binomială este aproximată printr-o repartiție Poisson.

S-a arătat că eroarea standard depinde numai de proporția cazurilor favorabile și de volumul populației originare. Acest lucru nu poate fi privit drept un paradox dacă-l interpretăm în lumina premiselor care stau la baza unui sondaj simplu. Într-adevăr, după cum știm, sondajul simplu se întîlnește în practică în cazurile în care sau fiecare element al eșantionului este *reintrodus* în populație înaintea următoarei extrageri, sau populația originară este, efectiv ori practic, infinită. În ambele cazuri populația este inepuizabilă și nici un eșantion nu este mai în măsură decît altul să includă toate elementele populației.

Mai trebuie subliniat un alt fapt, și anume acela că eroarea standard a proporției de cazuri favorabile variază invers proporțional cu $\sqrt{n}$; prin urmare, cînd n crește, eroarea standard descrește proporțional cu $\sqrt{n}$. Dacă, de pildă, un eșantion de 100 unități ne dă o eroare standard de 10%, pentru a o reduce la jumătate trebuie să luăm un eșantion de patru ori mai mare ($n = 400$).

În sfîrșit, încă o remarcă: eroarea standard poate fi utilizată pentru a măsura destul de bine imprecizia unei estimații a lui p; cu cît fluctuațiile de sondaj sînt mai mari, cu atît eroarea standard este mai mare, deși valoarea teoretică p a proporției este mereu aceeași. Inversul erorii standard $1/s'_p$ sau un multiplu al său, ales convenabil, poate fi considerat drept indicator de măsură a *preciziei*. Așadar, *precizia proporției observate variază direct proporțional cu rădăcina pătratică a volumului eșantionului.*

5.10.3 Compararea a două proporții de sondaj ale unei caracteristici-atribut oferite de două eșantioane diferite

Să presupunem că două eșantioane luate din populații diferite dau proporțiile p_1 și p_2 de elemente posedînd o anumită caracteristică A, volumele respective ale eșantioanelor fiind n_1 și n_2.

Scanned with OKEN Scanner

În cadrul acestor sondaje, ne putem afla în unul din următoarele trei cazuri diferite de analiză a informațiilor oferite de cele două eșantioane.

1. Nu avem nici o indicație teoretică în ceea ce privește proporția de elemente A în populațiile din care s-a extras unul sau altul dintre eșantioane însă putem admite ipoteza de lucru că cele două populații sînt similare în ceea ce privește proporția acestor elemente; de pildă, putem admite că această proporție este dată de media ponderată p_0 a proporțiilor de sondaj în cele două eșantioane luate împreună -- acesta ar fi cel mai bun indicator de care am putea dispune:

$$p_0 = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}. \tag{5.52}$$

Folosind această ipoteză ne putem întreba dacă cele două proporții p_1 și p_2 diferă semnificativ de p_0.

Erorile standard s'_{p_1} și s'_{p_2}, în cele două eșantioane, sînt:

$$s'_{p_1} = \sqrt{\frac{p_0 q_0}{n_1}}; \quad s'_{p_2} = \sqrt{\frac{p_0 q_0}{n_2}}.$$

Întrucît sîntem în condițiile unor sondaje simple, media diferenței între p_1 și p_2 va fi egală cu zero, eroarea standard a diferenței $s'_{p_1} - s'_{p_2}$ — eșantioanele fiind independente — va fi definită de relația:

$$s'_{p_1 - p_2} = \sqrt{p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \tag{5.53}$$

Dacă valoarea observată a diferenței dintre proporții va fi mai mare decît întreitul erorii standard a diferenței calculate potrivit relației (5.53), atunci, evident, ea trebuie interpretată statistic, ca fiind semnificativă; în caz contrar, diferența observată trebuie atribuită fluctuațiilor de sondaj.

2°. Proporțiile de elemente A nu sînt egale în cele două populații, p_1 și p_2 fiind valorile lor teoretice. În acest caz erorile standard de sondaj ale proporțiilor în cele două eșantioane sînt:

$$s'_{p_1} = \sqrt{\frac{p_1 q_1}{n_1}}; \quad s'_{p_2} = \sqrt{\frac{p_2 q_2}{n_2}},$$

iar eroarea standard a diferenței este dată de relația:

$$s'_{p_1 - p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}. \tag{5.54}$$

Problema care se pune în acest caz este aceea de a ști dacă diferența dintre p_1 și p_2, este nesemnificativă în raport cu fluctuațiile de sondaj. Este evident că dacă diferența dintre proporțiile observate nu depășește întreitul valorii calculate pe baza relației (5.54), atunci este posibil ca ea să se anuleze în alte eșantioane, datorită fluctuațiilor de sondaj.

3°. Ca și în cazul precedent, proporțiile teoretice de elemente A sînt p_1 și p_2. De data aceasta însă în loc de a compara proporția p_1 cu p_2, o comparăm cu proporția elementelor A în cele două eșantioane luate împreună, adică cu media p_0:

$$p_0 = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

Scanned with OKEN Scanner

Se pune întrebarea de a şti în ce caz diferenţa între p_1 şi p_0 poate apare drept efect al fluctuaţiilor de sondaj, p_0 fiind proporţia reală de elemente A în cele două eşantioane întrunite.

Răspunsul la întrebarea pusă este acelaşi ca şi în cazul 2 de mai sus; calculul este însă puţin mai complicat datorită faptului că erorile standard ale lui p_1 şi p_2 sînt dependente.

Se poate arăta că eroarea standard a diferenţei dintre p_1 şi p_0, notată cu $s'_{p_0-p_1}$, este definită de următoarea relaţie:

$$s'_{p_0-p_1} = \sqrt{\frac{p_0 q_0}{n_1 + n_2} \cdot \frac{n_2}{n_1}}. \qquad (5.55)$$

Dacă diferenţa dintre p_0 şi p_1 nu depăşeşte întreitul valorii $s'_{p_0-p_1}$, ea poate fi considerată drept efect al fluctuaţiilor de sondaj.

Aplicaţia 13. Această aplicaţie este o variantă asociată a aplicaţiei 11 de mai sus.

În afară de eşantionul celor 900 de oi ales în judeţul A se mai extrage încă un alt eşantion, din judeţul B, format din 3600 de oi. În al doilea eşantion se constată un număr de 864 de oi merinos, reprezentînd deci o proporţie de 0,24 sau 24% din volumul eşantionului.

Se pune problema să se compare proporţia de 20% a oilor merinos din eşantionul extras în judeţul A cu proporţia oilor merinos din ambele eşantioane, luate laolaltă şi care reprezintă $\frac{1044}{4200} = 0{,}2320$ sau 23,2%.

După cum se vede, diferenţa este de 3,2%.

Să examinăm dacă această diferenţă observată este semnificativă sau poate fi atribuită fluctuaţiilor de sondaj; în acest scop vom determina eroarea standard a diferenţei pe baza relaţiei (5.55):

$$s'_{p_0-p_1} = \sqrt{\frac{p_0 q_0}{n_1 + n_2} \cdot \frac{n_2}{n_1}} = \sqrt{\frac{0{,}232 \cdot 0{,}768}{3\,600 + 900} \cdot \frac{3\,600}{900}} \simeq 0{,}0125$$

sau 1,25%.

Diferenţa observată de 3,2% nu depăşeşte întreitul erorii standard a diferenţei de sondaj dintre cele două proporţii, astfel încît ea poate fi atribuită fluctuaţiilor de sondaj.

Cît priveşte nivelul de semnificaţie probabilistică cu care tragem această concluzie, el poate fi aproximat şi cu ajutorul repartiţiei normale, deşi, în speţă, proporţiile cazurilor favorabile urmează legi de repartiţie binomiale. Acest lucru este posibil datorită faptului că volumele eşantioanelor sînt mari. Vom considera deci repartiţiile proporţiilor drept normale şi vom alege — într-o primă necesitate practică — un nivel de semnificaţie, să zicem, de 1%.

Pentru a interpreta rezultatul obţinut în limita acestui nivel de semnificaţie este necesar, după cum ştim, calculul mărimii d, aşa cum am procedat şi în cazul variabilelor cantitative:

$$d = \frac{p_0 - p_1}{s'_{p_0-p_1}}. \qquad (5.56)$$

În exemplul examinat avem:

$$d = \frac{0{,}232 - 0{,}2}{0{,}0125} = 2{,}56,$$

mărime ceva mai mică decît cea teoretică, 2,58, corespunzătoare nivelului de semnificaţie ales. Sîntem într-un caz interesant în care mărimea observată d este cuprinsă între 1,96 şi 2,58, caz în care ipoteza diferenţei nule nu se acceptă, dar nici nu se respinge, fiind considerată de semnificaţie îndoielnică.

Este evident că la un nivel de semnificaţie mai puternic, să zicem de 0,2%, căruia îi corespunde o mărime teoretică d egală cu 3, rezultatul nostru ar trebui să fie interpretat drept *nesemnificativ* şi, în consecinţă, ar trebui admisă ipoteza diferenţei nule, trăgîndu-se în final concluzia că diferenţa observată se datorează fluctuaţiilor de sondaj cu o probabilitate de 0,998 sau 99,8%.

5.10.4 Abateri de la condiţiile sondajului simplu şi efectul lor asupra erorii standard a proporţiei cazurilor favorabile

1°. Un prim tip de abatere de la condiţiile unui sondaj simplu în cazul unei caracteristici-atribut ar fi acela în care proporţia elementelor A, în fiecare din eşantioanele extrase, n-ar rămîne constantă; o asemenea abatere ar avea loc în cazul cînd luăm eşantioane dintr-o populaţie variabilă în spaţiu şi timp sau din straturi diferite ale unei populaţii neomogene.

De pildă, ne putem imagina următorul proces de sondaj:

În fiecare lună a anului extragem, la întîmplare, eşantioane de cîte 100 persoane din fiecare judeţ şi înregistrăm de fiecare dată numărul absolut şi, respectiv, proporţia celor cu studii economice superioare.

Este evident că în acest caz pot exista diferenţe esenţiale de la o lună la alta, ca şi de la un judeţ la altul, în ceea ce priveşte proporţia persoanelor cu studii economice superioare, diferenţe generate, să spunem, de mobilitatea spaţială a cadrelor. Tocmai din acest motiv relaţiile (5.49) şi (5.51), definind erorile standard ale numărului şi proporţiei cazurilor favorabile, nu mai sînt aplicabile în cazul unui astfel de sondaj. Este aplicabilă însă următoarea relaţie pentru proporţia cazurilor favorabile:

$$s'_p = \sqrt{\frac{p_0 q_0}{n} + \frac{n-1}{n} \cdot s_p^2}. \tag{5.57}$$

2°. Un alt tip de abatere de la condiţiile sondajului simplu are loc atunci cînd proporţia elementelor posedînd o caracteristică-atribut nu rămîne constantă la extragerea fiecărui element al eşantionului. În acest caz, se poate folosi următoarea formulă pentru eroarea standard a numărului de cazuri favorabile:

$$s'_{np} = \sqrt{np_0q_0 - ns_p^2}. \tag{5.58}$$

Împărţind termenii sumei de sub radical cu n^2, obţinem eroarea standard a proporţiei:

$$s'_p = \sqrt{\frac{p_0 q_0}{n} - \frac{s_p^2}{n}}. \tag{5.59}$$

După cum se vede — făcînd comparaţie cu relaţiile (5.49) şi (5.51) — erorile standard în acest caz sînt mai mici.

3°. Vom presupune acum că cele două tipuri de abateri de la condiţiile sondajului simplu arătate mai sus nu au loc; în schimb, are loc o oarecare

Scanned with OKEN Scanner

dependență dintre evenimente, ceea ce, desigur, relevă un nou tip de abatere de la condițiile sondajului simplu.

Se poate arăta că în acest caz dispersia numărului de cazuri favorabile nu mai este npq, ci dată de următoarea relație:

$$s^2 = npq + 2pq(R_{12} + R_{13} + \ldots + R_{23} + \ldots),$$

în care R_{12}, R_{13} etc. sînt coeficienții de corelație dintre rezultatele primului și celui de-al doilea eveniment, dintre rezultatele primului și celui de-al treilea etc.; numărul lor este egal cu $\frac{n(n-1)}{2}$.

Dacă media lor aritmetică este R, atunci dispersia numărului de cazuri favorabile o putem scrie astfel:

$$s^2_{np} = npq[1 - R(n-1)],$$

iar eroarea standard:

$$s'_{np} = \sqrt{npq[1 + R(n-1)]}. \tag{5.60}$$

În mod corespunzător, eroarea standard a *proporției* cazurilor favorabile, simplificînd prin n^2, este:

$$s'_p = \sqrt{\frac{pq}{n}[1 + R(n-1)]}. \tag{5.61}$$

Este evident că R este un coeficient de corelație pentru toate perechile posibile de rezultate în cele n evenimente ale fiecărui eșantion. Dacă acest coeficient este pozitiv, erorile standard (5.60) și (5.61) vor *depăși* pe cele definite de (5.49) și respectiv (5.51). În cazul sondajului simplu și invers, ele vor fi mai mici, dacă coeficientul de corelație R este negativ.

Este de observat că în cazul unui coeficient de corelație pozitiv sîntem, de fapt, în cazul tipului de abatere de la condițiile sondajului simplu descris la punctul 1° de mai sus. Dacă coeficientul de corelație este negativ sîntem în cazul tipului de abatere de la condițiile sondajului simplu descris la punctul precedent.

Aplicația 14. O ilustrare simplă a dependenței evenimentelor oferă cazul extragerii succesive a n bile dintr-o urnă, conținînd un număr finit de N bile de două culori, să zicem albe și negre, cu proporțiile respective p și q, după fiecare extracție bila *nerevenind* în urnă.

În acest caz se poate demonstra că eroarea standard a numărului mediu de cazuri favorabile și cea a proporției acestor cazuri sînt, respectiv, următoarele:

$$s'_{np} = \sqrt{npq\left(1 - \frac{n-1}{N-1}\right)} = \sqrt{npq\,\frac{N-n}{N-1}} \tag{5.62}$$

$$s'_p = \sqrt{\frac{pq}{n}\left(1 - \frac{n-1}{N-1}\right)} = \sqrt{\frac{pq}{n}\cdot\frac{N-n}{N-1}}. \tag{5.63}$$

Aceste rezultate sînt deosebit de importante și au largi aplicații practice.

Scanned with OKEN Scanner

5.11 TEHNICI PARTICULARE DE SONDAJ

În acest paragraf vom prezenta cîteva tehnici particulare de elaborare și organizare a planului de sondaj.

5.11.1 Sondaj de serii

În numeroase cercetări aplicative sînt cazuri în care unitățile de observare ale populației originare formează în mod natural niște *ansambluri*, de obicei neomogene și care în termenii teoriei sondajului pot fi considerate drept unități *complexe* de observare. În categoria unor asemenea unități pot fi incluse, de pildă, echipele de muncitori din diferite exploatări (miniere, forestiere etc.), brigăzile de la diferite întreprinderi industriale sau C.A.P. ș.a.m.d.

Cînd cercetăm, pe bază de sondaj, populații statistice formate din astfel de unități complexe, este evident că problema care se pune este aceea de a fi incluse în eșantion, cu respectarea principiilor fundamentale ale sondajului, tocmai aceste unități complexe sau *serii* de unități elementare. Efectuăm astfel un *sondaj de serii*. În acest fel de sondaj, o serie este asimilată cu o unitate simplă din celelalte tipuri de sondaj, iar o caracteristică studiată apare sub forma nivelului ei mediu în seria i, adică $\bar{x}_i$.

În aceste condiții toate unitățile simple din care se compune seria extrasă aleator sînt supuse înregistrării, fără nici o excepție, pentru ca nivelul mediu caracterizînd o serie oarecare supusă cercetării să fie cît mai real.

Dacă notăm cu r volumul eșantionului aleator (adică un anumit număr de serii extrase la întîmplare), atunci media de sondaj a caracteristicii studiate o definim potrivit modelului general cunoscut:

$$M(\bar{x}_i) = \frac{\sum_{i=1}^{r} \bar{x}_i}{r}. \tag{5.64}$$

De asemenea, dispersia mediei de sondaj notată cu $\bar{\delta}^2_{\bar{x}_i}$ este dată, prin definiție, de următoarea expresie:

$$D^2[M(\bar{x}_i)] = D^2_{M(\bar{x}_i)} = \bar{\delta}^2_{\bar{x}} = \frac{\sum_{i=1}^{r} [\bar{x} - M(\bar{x}_i)]^2}{r}. \tag{5.65}$$

După cum se vede, prin forma ei, dispersia mediei de sondaj în serii este identică cu cea de la sondajul simplu însă, prin conținutul ei, este evident de tipul unei dispersii *dintre* grupe.

Din punct de vedere organizatoric, sondajul de serii este mai convenabil și de aceea practicienii sînt înclinați a-i da preferință. Este de subliniat, totuși, că un astfel de sondaj duce la o eroare standard a mediei de sondaj mai mare decît cea pe care ne-o dă un sondaj simplu, individual. De aceea, este recomandabil ca volumul eșantionului să fie cît mai mare cu putință.

Expresiile de mai sus (5.64) și (5.65) privind media și dispersia în cazul sondajului de serii se referă la ipoteza potrivit căreia seriile sînt formate din numere egale de unități elementare și, în consecință, comportă un aparat matematic mai simplu.

Scanned with OKEN Scanner

Prin analogie cu logica construcțiilor anterioare privind erorile standard, vom defini și pentru sondajul de serii eroarea standard a mediei în cazul unei caracteristici *cantitative*, considerînd sondajul repetat ca și cel nerepetat.

a) în cazul unui sondaj *repetat*, eroarea standard a mediei este:

$$s'_{M(\bar{x}_i)} = \sqrt{\frac{\overline{\delta_{\bar{x}}^2}}{r}}; \tag{5.66}$$

b) în cazul unui sondaj *nerepetat*:

$$s'_{M(\bar{x}_i)} = \sqrt{\frac{\overline{\delta_{\bar{x}}^2}}{r}\left(1 - \frac{r}{R}\right)}, \tag{5.67}$$

în care R reprezintă numărul unităților complexe (seriilor) în populația originară.

Pentru o *caracteristică-atribut*, erorile standard respective ale mediei de sondaj sînt, prin definiție:

a) în cazul sondajului *repetat*:

$$s'_{M(\bar{x}_i)} = \sqrt{\frac{\overline{\delta_p^2}}{r}}; \tag{5.68}$$

b) în cazul sondajului *nerepetat*:

$$s'_{M(\bar{x}_i)} = \sqrt{\frac{\overline{\delta_p^2}}{r}\left(1 - \frac{r}{R}\right)}. \tag{5.69}$$

Observația 1. În cercetări aplicative este mai utilă folosirea coeficientului $\frac{R-r}{R-1}$ decît a celui dat în formulele de mai sus privind sondajul de serii nerepetat, întrucît are proprietatea de a micșora — foarte puțin însă — eroarea de sondaj.

Observația 2. Sondajul de serii este cu atît mai precis cu cît, în cadrul fiecărei serii, caracteristica studiată are variații mai mici; în particular, cînd dispersia pe serii este aceeași, sondajul devine absolut precis dar, în acest caz, organizarea lui nu mai prezintă nici un fel de avantaje.

În domeniul fenomenelor economice, unde lipsa de uniformitate în repartiția caracteristicilor este mare, sondajul de serii este de regulă mai puțin precis decît sondajul unităților individuale. El are avantajul de a putea fi mai ușor organizat.

În afară de sondajul de serii în forma arătată se mai poate folosi o formă *combinată* a acestuia, în care, într-o primă fază, se alege un număr anumit de serii, iar în faza a doua, din cadrul fiecărei serii, un anumit număr de unități elementare, cu respectarea, în ambele faze, a principiului fundamental al sondajului, garantînd fiecarei unități — complexe sau elementare — aceeași șansă de a fi inclusă în eșantion. De pildă, un asemenea sondaj combinat se folosește în prognoza recoltei medii la hectar, cînd într-o primă fază se alege un anumit număr de comune sau cooperative agricole de producție și abia în a doua fază din cadrul acestora, se alege un procent aleator de sectoare care se supun cercetării prin tehnici bine cunoscute.

Cititorul a putut deduce, desigur, că în cazul sondajului de serii egale combinat, expresiile matematice ce definesc erorile standard de sondaj ale

Scanned with OKEN Scanner

mediei sînt analoage celor pe care le-am examinat cu prilejul expunerii privind sondajul în două sau mai multe faze cu o singură caracteristică. Vom preciza şi acum că eroarea sondajului de serii, combinat, rezultă din suma erorilor celor două faze ale acestuia şi că este deci mai mare decît oricare dintre acestea.

5.11.2 Sondaj mecanic (sistematic)

O altă variantă particulară a sondajului simplu este aşa-anumitul *sondaj mecanic* sau *sistematic*, deşi ultimul termen nu este tocmai potrivit. Cînd elementele unei populaţii sînt aranjate într-o anumită ordine în spaţiu sau în timp, de pildă, ordinea în care apar persoanele înscrise alfabetic într-o carte de telefon, preţurile publicate regulat în fiecare săptămînă etc., este convenabil uneori ca eşantionul aleator să se formeze alegîndu-se fiecare element la intervale egale într-un anumit sens. Un astfel de interval poartă denumirea de *pas de numărare*. Un pas de numărare se poate obţine, de pildă, împărţind volumul populaţiei originare la volumul eşantionului, stabilit judicios, în funcţie de precizia şi probabilitatea cu care vrem să obţinem rezultatele cercetării.

Acest procedeu de sondaj poartă denumirea de *sondaj mecanic*.

Este evident că eficienţa şi caracterul sondajului mecanic depind în mod hotărîtor de structura populaţiei originare. Din acest unghi de vedere sondajul mecanic poate fi asimilat unui sondaj tipic-stratificat în care, din fiecare grupă tipică, se alege cîte o singură unitate. El diferă însă de un sondaj stratificat prin aceea că grupele în care se divide populaţia nu sînt tipice în adevăratul înţeles al cuvîntului deşi, într-o oarecare măsură, ele sînt omogene.

Dacă ordinea în care sînt aşezate unităţile populaţiei originare este întîmplătoare, atunci toate tipurile, prin structura lor, vor fi asemănătoare, abstracţie făcînd, desigur, de deosebiri cu totul întîmplătoare. În acest caz, prin esenţa lui, sondajul mecanic nu este decît un sondaj simplu şi în consecinţă calculul erorilor de sondaj se va face ca şi în cazul acestuia din urmă. De precizat că se aplică, de obicei, formulele de calcul ale erorilor stabilite pentru sondaj simplu repetat. Folosind aceste formule mărim precizia sondajului, întrucît erorile date de sondajul mecanic sînt, în general, mai mici decît ale sondajului simplu repetat, însă ceva mai mari decît ale sondajului simplu nerepetat.

Cînd ordinea unităţilor populaţiei originare nu este cu totul întîmplătoare, sondajul mecanic încetează de a mai fi întîmplător; în aceste condiţii, sondajul conţinînd cîte o unitate din fiecare tip, va fi mai preferabil sondajului simplu.

Sondajul mecanic poate da rezultate cu totul nesatisfăcătoare — deşi în cazul cînd populaţia este eterogenă el pare să evite alegerea în eşantion a unui număr prea mare de elemente nereprezentative — sau, cel puţin, rezultate nesigure dacă elementele populaţiei originare sînt afectate de o periodicitate oarecare. Acest pericol poate deveni foarte grav dacă datele privind diferite elemente ale populaţiei originare se înregistrează în ordine cronologică. Într-adevăr, în acest caz pot interveni anumite schimbări ritmice în sînul populaţiei originare, aşa cum se întîmplă, de pildă, cu anumite serii în timp oscilatorii.

Cele arătate mai sus, sugerează în mod necesar concluzia că sondajul mecanic poate fi recomandat numai atunci cînd există motive temeinice de a afirma, apriori, că intervalele dintre unităţile populaţiei, incluse în eşantion, nu sînt afectate de nici o proprietate sistematică a populaţiei cercetate.

Scanned with OKEN Scanner

5.11.3 Sondaj secvențial

Un loc aparte în teoria sondajului îl ocupă sondajul *secvențial*. Tehnica lui are drept scop principal formarea unui eșantion, de dimensiuni strict necesare — eșantion de maximă economicitate — în măsură să ne furnizeze totuși un volum de informații suficient pentru luarea unei decizii corecte, cu un anumit grad de siguranță prestabilită. În procesul sondajului secvențial tehnica de formare a unui astfel de eșantion urmează — într-un anumit sens — o serie de etape de lucru inverse acelora folosite, de obicei, la celelalte tipuri de sondaj.

Într-adevăr, încă în momentul inițial al procesului de desfășurare a acestui tip de sondaj, cercetătorului i se pune problema de a decide — pe baza unui grad de siguranță fixat — care vor fi *limitele* între care va putea sau nu va putea fi acceptată estimația dată de eșantion ca fiind concordonată cu valorile prescrise ale parametrului studiat. În continuare, după ce limitele de acceptabilitate au fost fixate, sînt prelevate secvențial — unul cîte unul sau în grupe — elementele eșantionului. Dacă într-un stadiu al prelevării, calculul și analiza indicatorilor de estimație ne indică o depășire a limitelor proiectate corespunzînd volumului atins al eșantionului, atunci ipoteza că parametrul populației are valoarea prescrisă se respinge.

După cum se vede, în cursul procesului de sondaj, cercetătorul analizează secvențial o întreagă succesiune de eșantioane, rînd pe rînd, după ce la eșantionul precedent se adaugă cîte o unitate elementară sau cîte o grupă de asemenea unități extrase aleator din populația originară, continuarea sau terminarea extragerii se decide după fiecare extragere, în funcție de rezultatul obținut pe baza observărilor făcute pînă în acel moment.

Dată fiind această particularitate a sondajului secvențial, volumul eșantionului nu poate fi fixat dinainte; el trebuie definit deci ca orice variabilă aleatoare.

Teoria sondajului secvențial s-a dezvoltat îngemănat cu teoria deciziilor și teoria verificării ipotezelor statistice.

Aplicația 15. La un oficiu de recepție al unui mare magazin alimentar sosește un lot de 4 000 cutii de conserve de carne. Se pune problema ca lotul să fie supus controlului statistic de calitate, caracteristica cercetată fiind gramajul; o cutie de conserve se consideră a fi corespunzătoare dacă greutatea sa nu este mai mică de 250 grame și necorespunzătoare în caz contrar. Potrivit unei înțelegeri prealabile între producătorul furnizor al cutiilor de conserve și beneficiar — magazinul alimentar — controlul trebuie efectuat prin sondaj secvențial, ținîndu-se seama de următorii indicatori:

$p_1 = 0{,}1$, reprezentînd nivelul de calitate acceptabil;
$\alpha_1 = 0{,}02$ — riscul furnizorului;
$p_2 = 0{,}3$ — nivelul de calitate tolerat;
$\beta_2 = 0{,}03$ — riscul beneficiarului;
d_n — numărul de cutii de conserve defecte, găsite pînă la a n-a cutie controlată;
a_n — numărul de acceptare;
r_n — numărul de respingere;

Scanned with OKEN Scanner

Dacă după cea de-a n-a probă obținem inegalitatea $a_n < d_n < r_n$, atunci controlăm și cea de-a $n + 1$ cutie ș.a.m.d.
Dacă avem:

$$d_n \leqslant a_n$$

oprim controlul și acceptăm lotul ca fiind corespunzător (acceptăm ipoteza $p = p_1$).

În sfîrșit, dacă la a n-a probă înregistrăm

$$d_n \geqslant a_n,$$

decizia ce o luăm este cea de respingere a întregului lot de cutii de conserve (adică, acceptăm ipoteza $p = p_2$).

Din teoria controlului statistic de recepție se știe că a_n și r_n se definesc după cum urmează

$$a_n = -h_1 + s \cdot n \tag{5.70}$$

$$r_n = h_2 + s \cdot n, \tag{5.71}$$

unde h_1, h_2 și s sînt dați de următoarele expresii:

$$h_1 = \frac{\log \frac{1-\alpha_1}{\beta_2}}{\log \frac{p_2}{p_1} - \log \frac{1-p_2}{1-p_1}}; \quad h_2 = \frac{\log \frac{1-\beta_2}{\alpha_1}}{\log \frac{p_2}{p_1} - \log \frac{1-p_2}{1-p_1}}; \quad s = \frac{\log \frac{1-p_1}{1-p_2}}{\log \frac{p_2}{p_1} - \log \frac{1-p_2}{1-p_1}}.$$

Făcînd înlocuiri cu datele exemplului, obținem:

$$h_1 = \frac{\log \frac{0{,}98}{0{,}03}}{\log \frac{0{,}3}{0{,}1} - \log \frac{0{,}70}{0{,}90}}; \quad h_2 = \frac{\log \frac{0{,}97}{0{,}02}}{\log 3 - \log 0{,}777}; \quad s = \frac{\log \frac{0{,}90}{0{,}70}}{\log 3 - \log 0{,}777}$$

$$h_1 = 2{,}539; \quad h_2 = 2{,}827; \quad s = 0{,}182$$

Deci, valorile a_n și r_n sînt:

$$a_n = -2{,}539 + 0{,}182 \cdot n$$

$$r_n = 2{,}827 + 0{,}182 \cdot n.$$

Dînd pe rînd lui n valorile 1, 2, ... obținem numerele de acceptare a_1, a_2, ... și cele de respingere r_1, r_2, ...

Scanned with OKEN Scanner

În tabelul 5.9 am trecut datele calculate.

Tabelul 5.9

Nr. de ordine al probei	Nr. de acceptare	Nr. de cutii de conserve defecte	Numărul de respingere
1	.	0	.
2	.	0	.
3	.	1	.
4	.	1	4
5	.	1	4
6	.	1	4
7	.	1	5
8	.	1	5
9	.	2	5
10	.	2	5
11	.	3	5
12	.	4	6
13	.	4	6
14	.	5	6
15	0	5	6
16	0	5	6
17	0	5	7
18	0	6	7
19	0	6	7
20	0	6	7
21	1	6	7
22	1	7	7
23	1		8
24	1		8
25	1		8
26	2		8
27	2		9
28	2		9
29	2		9
30	3		9

Verificarea s-a oprit la $n = 22$ deoarece pînă la a 22-a cutie s-au găsit 7 cutii rele și deci $d_{22} \geqslant r_{22}$.

Lotul trebuie deci respins deoarece numărul de defecte $d_{22} = 7$ este egal cu numărul de respingere $r_{22} = 7$.

Pînă la $n = 14$ nu se poate lua decizia de acceptare deoarece se obține $a_n < 0$, ceea ce este absurd. Pînă la $n = 3$ nu se poate lua decizia de respingere deoarece se obține $r_n > n$, ceea ce este de asemenea absurd.

5.12 DIMENSIONAREA EȘANTIOANELOR

În procesul organizării cercetărilor pe bază de sondaj o primă problemă care trebuie să fie rezolvată este cea privitoare la *eroarea de sondaj admisibilă* și *nivelul probabilității* cu care vor fi garantate rezultatele sondajului. Nivelele acestor doi indicatori apar deci ca fiind date înainte de începerea formării eșantionului respectiv. Prin urmare, necunoscut rămîne doar acel *volum minim* al eșantionului care ne va asigura exactitatea cerută.

Scanned with OKEN Scanner

Acest volum poate fi stabilit considerînd în calcul *eroarea limită de sondaj* pe care am definit-o în 5.4.5 ca produs $d \cdot s'$ (vezi (5.8)).

Dacă pentru acest produs introducem notaţia

$$\Delta = d \cdot s'_{\bar{x}}, \tag{5.72}$$

atunci dimensiunile eşantioanelor, pentru diferite tipuri de sondaj, pot fi uşor calculate. În tabelul 5.10 sînt date volumele eşantioanelor în cazul unui sondaj pur aleator, pentru cele două modele fundamentale de sondaj — repetat şi nerepetat — vizînd atît caracteristici cantitative, cît şi caracteristici-atribut.

Tabelul 5.10

DIMENSIUNILE EŞANTIOANELOR ÎN CAZUL SONDAJELOR PUR ALEATOARE

Modelul sondajului preconizat	Formulele ce se folosesc pentru:		
	Valoarea medie	Nivelul proporţiei în cazul caracteristicilor-atribut	Nivelul proporţiei chiar dacă ea nu poate fi aproximată
a) repetat	$n = \frac{d^2 \cdot \sigma^2}{\Delta^2}$	$n = \frac{d^2 w(1-w)}{\Delta^2}$	$n = \frac{0{,}25 \cdot d^2}{\Delta^2}$
b) nerepetat	$n = \frac{d^2 \cdot \sigma^2 \cdot N}{N \cdot \Delta^2 + d^2 \cdot \sigma^2}$	$n = \frac{d^2 \cdot N \cdot w(1-w)}{N \cdot \Delta^2 + d^2 w(1-w)}$	$n = \frac{0{,}25 \cdot d^2 N}{N\Delta^2 + 0{,}25 d^2}$

În cazul sondajului tipic — stratificat — am putut sesiza în paragraful (5.8) modul în care se poate determina volumul eşantionului.

Vom preciza cu acest prilej că în cazul unui sondaj proporţional cu volumul fiecărui strat (N_j), volumul eşantionului ce se extrage din fiecare grupă tipică (n_j) este dat de următoarea formulă

$$n_j = n \cdot \frac{N_j}{N}, \tag{5.73}$$

în care indicatorii folosiţi ne sînt de acum cunoscuţi:

n — volumul global al eşantionului;
N_j — volumul grupei tipice, cu numărul de ordine j;
N — volumul populaţiei originare.

În ipoteza unui sondaj tipic care ţine seama şi de *variabilitatea caracteristicii* (5.37) oferind cea mai mică eroare de sondaj, procentul eşantionului în fiecare grupă tipică trebuie să fie proporţional cu abaterea medie pătratică a grupei respective. Se folosesc următoarele formule pentru calculul eşantioanelor:

$$n_j = \frac{n N_j s_j}{\sum_{j=1} N_j s_j} \text{ (pentru valoarea medie);} \tag{5.74}$$

$$n_j = \frac{n \cdot N_j \sqrt{w_j(1-w_j)}}{\sum_{j=1} N_j \sqrt{w_j(1-w_j)}} \text{ (pentru proporţie).} \tag{5.75}$$

Scanned with OKEN Scanner

Nu este greu de observat că în cazul unui sondaj de serii, numărul necesar de serii prelevate în eșantion poate fi determinat — ținînd seama de natura acestui sondaj — folosindu-se aceleași formule ca la sondajul pur aleator, cu deosebirea că în loc de indicatori simbolizați prin N, n și s^2 vor fi folosiți R, r, $\bar{\delta}^2$, ultimul indicator fiind dispersia interserii.

În încheiere vom sublinia că de volumul eșantionului (n) — și aceasta în mod exclusiv — depinde gradul de apropiere dintre media de sondaj și media populației originare, dispersia fiind aceeași. Greutatea specifică a eșantionului nu are nici o influență în această privință. Așa, de pildă, volumul eșantionului, constituind 1% din populația de 100 000 unități, ne oferă o eroare de sondaj mai mică decît un eșantion de 2% dintr-o populație originară de 1 000 unități, dispersia în cele două populații fiind aceeași. După cum se vede, dimensionarea corectă a eșantioanelor constituie o problemă-cheie a sondajului statistic.

Scanned with OKEN Scanner

CAPITOLUL 6

Sondajul statistic (II). Sondaje pe bază de eșantioane mici. Concluzii

6.1 CONSIDERAȚII INTRODUCTIVE

Sintetizînd problematica capitolului precedent, rezervat sondajelor pe bază de eșantioane mari, putem reține două premise esențiale pe care se fundamentează aceste tipuri de sondaj. Una din ele justifică afirmația că, în general, repartițiile de sondaj ale diferitelor valori tipice sînt aproximativ normale sau cel puțin unimodale, iar cealaltă — că valorile tipice, calculate pe bază de eșantioane mari, sînt suficient de apropiate de valorile corespunzătoare ale populației originare. De aceea, valorile tipice ale repartițiilor de sondaj pot servi drept indicatori de estimație, fiind posibilă, totodată, cunoașterea preciziei estimațiilor realizate.

În teoria sondajelor bazate pe eșantioane de volum redus premisele de mai sus nu mai sînt valabile și, în consecință, sînt introduse, în calcul, în mod natural, restricții destul de severe, care antrenează și o tehnică nouă de tratare a problemelor de sondaj.

Sondajele bazate pe eșantioane de volum redus au largi aplicații în cercetările economice. Adeseori, în producție, economiștii sînt nevoiți să ia decizii importante, pe baza unor date sumare, oferite de eșantioane mici. Asemenea decizii, chiar dacă sînt afectate de o anumită „nedeterminare", pot avea o bază științifică, dacă se recurge la anumite metode ale statisticii matematice, special elaborate.

În mod firesc a apărut necesitatea efectuării de sondaje, în care numărul unităților să nu depășească valorile de ordinul 10—20. Aceste sondaje se consideră a fi de *volum redus* și se folosesc, de pildă, la stabilirea normelor tehnice de muncă. Practic, consumul specific de timp de muncă, pe operații, se înregistrează numai la un număr mic de muncitori, rezultatele observate extinzîndu-se apoi asupra întregii populații de muncitori, pe bază de estimații.

6.2 ESTIMAREA VALORILOR TIPICE ALE POPULAȚIEI ORIGINARE PE BAZĂ DE EȘANTIOANE MICI

Estimarea mediei. Vom preciza, din capul locului, că o expunere completă a bazelor teoretice ale estimării parametrilor populației originare, în cazul eșantioanelor mici, ar implica tratarea mai amplă a unor probleme matematice

Scanned with OKEN Scanner

adeseori de mare complexitate. De aceea, în cele ce urmează, vom examina doar cazurile particulare de estimare a unor parametri de bază, cum sînt media și abaterea medie pătratică (sau dispersia).

Dacă avem un număr redus de n valori aleatoare de sondaj $x_1, x_2, \ldots, x_n$ ($n < 30$), media de sondaj, definită de relația

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{6.1}$$

poate fi luată drept indicator de estimare a mediei m în populația originară din care a fost extras eșantionul. Acest procedeu de estimare se folosește în practică și se poate arăta că el are o deplină justificare teoretică.

După cum știm, în cazul unor eșantioane — indiferent dacă sînt mici sau mari — extrase într-o populație normală, estimarea mediei pe baza relației (6.1) poate fi considerată ca cea mai bună, în sensul că dispersia ei este mai mică decît cea pe care o putem obține cu ajutorul oricărui alt procedeu de estimare a acestei valori tipice.

Prin urmare, pentru estimarea mediei, teoria și practica sînt aceleași, atît pentru eșantioane mari, cît și pentru cele mici.

Estimarea dispersiei. Vom nota, ca și pînă acum, cu σ^2 și m, dispersia și, respectiv, media în populația originară.

Dacă m, valoarea medie, este cunoscută, atunci pentru estimarea dispersiei ce caracterizează populația originară putem lua dispersia datelor eșantionului în raport cu m: cu alte cuvinte, indicatorul de estimație, pe care-l notăm cu s^2, ar fi definit de relația:

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2. \tag{6.2}$$

Dar, valoarea lui m, în general, nu ne este cunoscută, astfel încît ea însăși trebuie estimată.

Se poate demonstra că dispersia valorilor de sondaj nu poate fi acceptată drept indicator de estimație a dispersiei în populația originară, întrucît aceasta ar echivala cu a introduce, conștient, o eroare de sondaj sistematică de mărimea $(\bar{x} - m)^2$. Acest termen definește chiar eroarea de sondaj a mediei care, după cum știm, este egală cu $\frac{\sigma^2}{n}$. Prin urmare, pe această bază putem construi un nou indicator de estimare a dispersiei în populația originară pe care-l vom nota cu $s^2_{\sigma^2}$.

Avem deci:

$$s^2_{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) + \frac{\sigma^2}{n}.$$

Întrucît valoarea lui σ^2 este de obicei necunoscută, o putem aproxima cu s^2. Procedînd astfel, obținem:

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 + \frac{1}{n} \cdot s^2$$

Scanned with OKEN Scanner

sau

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \qquad (6.3)$$

Acceptarea acestui indicator drept indicator de estimație a dispersiei în populația originară va permite eliminarea erorii sistematice de sondaj de care am amintit mai înainte.

6.3 CÎTEVA OBSERVAȚII PRIVITOARE LA ESTIMARE ȘI GRADELE EI DE LIBERTATE

1°. O primă observație: dacă, prin folosirea indicatorului (6.3), se vor face mai multe estimații pe bază de eșantioane mici, ele vor alcătui o *repartiție de sondaj* a dispersiei, avînd valoarea medie localizată, foarte probabil, lîngă valoarea adevărată a dispersiei în populația originară.

2°. Raționamentul pe baza căruia se deduce relația (6.3) nu face apel la gradul de mărime a lui n și de aceea, riguros vorbind, această relație poate fi aplicată și eșantioanelor mari. Este de observat totuși că dacă n este mai mare, valorile n și $n-1$ pot fi considerate practic ca fiind egale. În cazul eșantioanelor mari, luînd în calculul dispersiei $n-1$ în loc de n, obținem o diferență foarte mică. De aceea, în lucrările practice, efortul de a face rezultatele mai precise apare lipsit de sens.

3°. Mai remarcăm, în sfîrșit, că folosirea lui $\tilde{s}^2$ drept indicator de estimație „corectat", ținîndu-se seama de termenul $\frac{\sigma}{\sqrt{n}}$ este valabilă numai *în medie*. Aceasta ne obligă de a nu pierde din vedere faptul că dacă într-un eșantion aleator obținem întîmplător o dispersie de sondaj mai mare decît parametrul respectiv al populației originare, corecția înrăutățește situația și ne poate duce la rezultate aberante, inacceptabile sau chiar absurde.

4°. Vom da acum cîteva elemente explicative suplimentare pentru conceptul *grade de libertate* ale unei estimații. Cu noțiunea de grad de libertate, cititorul s-a mai întîlnit cu prilejul expunerii privind repartiția continuă χ^2 și repartiția Student. Cu acel prilej, s-a reținut, de pildă, că probabilitatea P ca să obținem pentru χ^2 o valoare mai mică sau mai mare decît un număr dat χ_0^2, adică

$$P(\chi^2 < \chi_0^2)$$

sau

$$P(\chi^2 > \chi_0)$$

variază în raport cu ν, *numărul gradelor* de libertate [1], definit de numărul grupelor — în cazul diferitelor serii statistice — a căror *frecvență poate fi fixată arbitrar*.

[1] Folosim — în acest context — un alt simbol pentru numărul gradelor de libertate ν, pentru a evita confuzia cu simbolul (s), folosit pentru abaterea medie pătratică de sondaj.

Scanned with OKEN Scanner

În general, dacă o variabilă aleatoare X este funcție de n variabile aleatoare, $X = f(X_1, X_2, ..., X_n)$ dintre care ν pot fi considerate *independente*, putem spune că X are ν grade de libertate. De obicei, numărul gradelor de libertate se află scăzînd din numărul variabilelor de care depinde X, numărul relațiilor prin care sînt legate aceste variabile. De pildă, dacă valorile unei caracteristici numerice sînt împărțite în n grupe, atunci aceasta are $\nu = n - 1$, grade de libertate, deoarece suma frecvențelor este cunoscută; într-adevăr, cunoașterea a $n - 1$ frecvențe din cele n frecvențe de grupă implică, evident, cunoașterea ultimei frecvențe.

Noțiunea gradelor de libertate se poate extinde în mod convenabil — și acest lucru este necesar — la estimarea parametrilor populației originare pe bază de eșantioane mici.

Dacă ne referim la estimațiile oferite de relațiile (6.1), (6.2) și (6.3), numărul gradelor de libertate va fi dat de numitorii acestora. Astfel, de pildă, estimația bazată pe relația (6.3) are numărul gradelor de libertate ν egal cu $n - 1$, întrucît în cantitatea $\sum(x_i - \bar{x})^2$ datele sondajului au mai fost folosite anterior pentru determinarea lui $\bar{x}$.

6.4 CRITERII ȘI TESTE DE APRECIERE A SEMNIFICAȚIEI STATISTICE A ESTIMAȚIILOR PE BAZĂ DE EȘANTIOANE MICI

6.4.1 Valoarea gnoseologică a testelor de semnificație

Expunînd problema criteriilor și testelor de apreciere a semnificației statistice a valorilor tipice de sondaj în cazul eșanticanelor mari, atrăgeam atenția că nivelul de încredere în oricare din acești indicatori ca indicatori de estimație a parametrilor respectivi ai populației originare depinde, în primul rînd, de stabilitatea lor, iar în al doilea rînd, de mărimea n a eșantionului.

Ce putem spune, în această privință, dacă ne referim la indicatorii de estimație bazați pe eșantioane mici?

Potrivit cercetărilor teoretice, se admite că și în cazul eșantioanelor mici stabilitatea indicatorilor de sondaj poate fi apreciată cu ajutorul indicatorilor împrăștierii, îndeosebi cu ajutorul abaterii mediei pătratice.

Cît privește eficiența conținutului informațional al eșantioanelor de volum redus, cercetările aplicative au arătat că ar fi greșit să se subaprecieze rolul pe care aceste eșantioane sînt chemate a-l îndeplini la cercetarea, pe bază de sondaj, a diferitelor populații. De pildă, ar fi greșit să se nege total sau parțial valoarea cognitivă pe care o au eșantioanele mici. Desigur, unele estimații bazate pe asemenea eșantioane vor fi mai bune decît altele, dar, de fapt — și aceasta cititorul trebuie să rețină în mod ferm —, nici una din aceste estimații nu va fi foarte precisă. În stadiul actual al cercetărilor teoretice, această afirmație nu comportă nici un fel de discuție, mai ales atunci cînd studiul pe bază de sondaj are drept obiect populații a căror normalitate este pusă la îndoială.

După cum am mai arătat, în procesul diferitelor cercetări aplicative, împrejurările de fapt cer să se facă deducții și să se ia decizii — desigur uneori cu caracter strict orientativ — chiar și pe baza unor date insuficiente. În astfel

Scanned with OKEN Scanner

de împrejurări putem încerca, adeseori cu succes, să localizăm valoarea parametrului în populația originară între anumite limite precise, practic utilizabile. Tocmai în această direcție s-a dezvoltat întreaga teorie a sondajelor pe bază de eșantioane mici. Generalizînd numeroase experimentări, ea oferă în prezent cercetătorului interesat criterii generale eficiente nu pentru a estima valoarea *exactă* a parametrului cercetat, ci pentru a preciza dacă valoarea observată a unui indicator de sondaj se abate sau nu, semnificativ, de la o anumită valoare dată în prealabil și de a clarifica, eventual, cauza unui asemenea fenomen.

Este evident că problema de a ști dacă abaterea se datorează fluctuațiilor inerente de sondaj sau diferitelor alte cauze — din care unele pot fi și cauze sistematice — prezintă, după cum știm, o deosebită importanță pentru statisticieni.

6.4.2 Ipoteza de normalitate. Aplicații ale repartiției Student

Precizăm, de la început, că testele și rezultatele teoretice, pe care le vom prezenta și analiza în continuare, menite să înlesnească studiul informațiilor oferite de eșantioane mici, se fundamentează pe ipoteza generală că populația originară din care s-a prelevat eșantionul are proprietățile unei populații normale, fără ca *proprietăți analoage să fie atribuite și datelor observate.*

Vom menționa însă că, în condițiile în care abaterile repartițiilor originare de la normalitate nu sînt prea mari, este de presupus că unele teste și rezultate teoretice de care ne ocupăm mai jos ar putea fi aplicate și în aceste condiții. Dealtfel, cercetările teoretice au ajuns să confirme această presupunere intuitivă.

Desigur, în procesul cercetărilor practice, putem avea motive temeinice să considerăm că populațiile originare supuse observării sînt semnificativ asimetrice, de pildă, ele pot fi în formă de U sau J. În aceste cazuri, metodele de analiză a eșantioanelor mici nu mai sînt aplicabile în condiții bune.

În teoria statisticii matematice s-a dovedit posibilă — și acest rezultat este remarcabil — construirea unor indicatori statistici ale căror repartiții de sondaj să fie complet independente de parametrii repartiției originare sau dependente doar de unii din acești parametri. Asupra acestor rezultate cititorul a fost informat în capitolul 3, cînd s-a expus repartiția lui Student.

Cu acel prilej am arătat — referitor la densitatea de repartiție a lui Student (3.65) că *ea nu depinde de abaterea medie pătratică a variabilei independente în raport cu care a fost definită.*

Această remarcabilă proprietate a repartiției Student o face să fie deosebit de utilă în cercetările statistice efectuate pe bază de sondaje mici. Într-adevăr, repartiția teoretică a lui t, fiind independentă de valoarea dispersiei în populația originară, are drept componente numai indicatorii calculați pe baza datelor de sondaj.

După cum știm, Student a introdus un indicator statistic avînd expresia (3.70).

Repartiția teoretică a acestui indicator a fost precis determinată, constituind o importantă contribuție a lui Student la dezvoltarea teoriei sondajului. Mai tîrziu, R. A. Fisher a propus o fundamentare teoretică mai riguroasă

Scanned with OKEN Scanner

pentru raportul (6.4) şi totodată i-a dat o formă nouă. Este vorba de expresia (3.70) pe care o reproducem şi aici:

$$t = \frac{\bar{x} - m}{s\sqrt{n}},$$

în care s, reprezintă rădăcina pătrată din expresia (6.2), iar ceilalţi indicatori ne sînt cunoscuţi.

Acest test urmează repartiţia teoretică a lui Student (3.65) cu $\nu = n - 1$ grade de libertate.

Dacă folosim abaterea medie pătratică de sondaj ca atare, deci rădăcina pătrată a expresiei (6.3), atunci testul lui Student capătă următoarea formă:

$$t = \frac{\bar{x} - m}{\tilde{s}} \cdot \sqrt{n - 1} \qquad (6.4)$$

O aplicaţie concretă ne va ilustra modul de folosire a testului t şi unele probleme deosebite de interpretare care se ridică cu acest prilej.

Aplicaţia 1. Să presupunem că ni se cere să apreciem, statistic, influenţa pe care o exercită asupra productivităţii muncii unui colectiv de muncitori introducerea unei maşini-unelte noi, de tip B, faţă de o maşină-unealtă mai veche, de tip A, aflată în funcţiune. Avem la dispoziţie, practic, un timp limitat în care trebuie să ne pronunţăm şi să decidem dacă este sau nu cazul ca maşina nouă să fie acceptată, apreciindu-se dacă sporul de productivitate pe care-l va aduce această maşină va compensa investiţiile importante care se fac pentru achiziţionarea ei.

Problema pusă se poate rezolva printr-o probă experimentală de producţie, în limita timpului disponibil, să zicem de 4—5 ore. Se aleg, în acest scop, de pildă, 10 muncitori cu aceeaşi pregătire tehnică, cu aceeaşi vechime în muncă şi, în general, cu aceeaşi îndemînare, care sînt supuşi la următoarea probă de producţie: fiecare lucrător lucrează cîte patru ore la noul tip de maşină-unealtă, timp în care se ţine evidenţa exactă asupra productivităţii orare medii a muncii pe care o realizează. Nivelul productivităţii orare medii a aceluiaşi colectiv de muncitori, la maşina-unealtă veche, de tip de A, este ştiut. Datele comparative se înregistrează în tabelul 6.1.

După cum se vede, maşina-unealtă de tip B pare să fie mai eficientă. În şapte cazuri din zece s-au înregistrat creşteri ale productivităţii muncii (s-au micşorat, în medie, numărul de minute necesare pentru executarea unei piese); cu toate acestea, în trei cazuri, maşina veche a dat rezultate mai bune (la muncitorii cu numerele de ordine 3, 6 şi 8).

În această situaţie se pune în mod firesc întrebarea dacă maşina-unealtă de tip mai nou (B) este într-adevăr mai eficientă.

Pentru a răspunde la această întrebare vom recurge la un procedeu tehnic specific de cercetare statistică care ne este de acum cunoscut. Vom face o ipoteză a diferenţei nule prin care presupunem că cele două maşini nu se deosebesc semnificativ, din punctul de vedere al randamentului de lucru pe care-l oferă. În esenţă, această ipoteză se reduce la a arăta că avantajul real al maşinii B faţă de maşina A este egal cu zero, adică valoarea medie a diferenţelor dintre productivităţile orare medii ale muncii înregistrate la maşina A şi maşina B este nulă.

Scanned with OKEN Scanner

Această ipoteză trebuie verificată, constatîndu-se statistic dacă media de selecție $\bar{x} = 0{,}4$ se abate semnificativ de la valoarea medie 0, presupusă ipotetic. În speță, avem un eșantion de volum redus ($n = 10$) și aceasta ne sugerează necesitatea aplicării testului t al lui Student, presupunîndu-se totodată că repartiția productivității muncii în populația originară este normală.

Tabelul 6.1

PRODUCTIVITATEA ORARĂ MEDIE A MUNCII UNUI COLECTIV DE 10 MUNCITORI ÎNREGISTRATĂ ÎN DECURS DE 4 ORE DE PRODUCȚIE LA DOUĂ TIPURI DE MAȘINI-UNELTE – A ȘI *B*

Numărul de ordine al muncitorului	Productivitatea orară medie la mașina tip A (minute pe o piesă produsă)	Productivitatea orară medie la mașina tip B (minute pe o piesă produsă)	Diferență (în minute pe o piesă produsă) între cele 2 productivități (x_i)
1	3,1	2,2	−0,9
2	3,0	2,0	−1,0
3	3,0	3,6	+0,6
4	6,2	5,3	−0,9
5	3,6	2,8	−0,8
6	4,0	4,2	+0,2
7	3,6	3,2	−0,4
8	2,8	3,2	+0,2
9	3,0	2,6	−0,4
10	6,0	5,4	−0,6
Total	3,83	3,43	−0,4

Să calculăm indicatorii de care avem nevoie, pe baza datelor din tabelul 6.1.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{-4}{10} = -0{,}4$$

$$\tilde{s}^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = \frac{1}{9}\left(4{,}38 - \frac{16}{40}\right) = \frac{2{,}78}{9};$$

$$\tilde{s}' = \sqrt{\frac{2{,}78}{9}} = \frac{1{,}66}{3} = 0{,}553;$$

$$D^2(\bar{x}) = \frac{\tilde{s}^2}{n} = \frac{2{,}78}{10}; \quad D(\bar{x}) = \frac{\tilde{s}'}{\sqrt{n}} = \frac{0{,}553}{3{,}16} = 0{,}175.$$

Și acum să determinăm valoarea de sondaj a mărimii t:

$$t = \frac{\bar{x} - m}{\tilde{s}'/\sqrt{n}} = \frac{-0{,}4 - 0{,}0}{0{,}175} = -2{,}28.$$

Din tabelele speciale cu valorile repartiției lui Student constatăm că pentru a obține o diferență semnificativă, în cazul unui nivel de semnificație de 1% cu 9 grade de libertate, t trebuie să fie egal sau mai mare decît numărul 3,250. În cazul nostru, t fiind mai mic decît valoarea sa teoretică putem trage

Scanned with OKEN Scanner

concluzia — cu un nivel de semnificație de 1% — că, din punct de vedere statistic, introducerea mașinii-unelte noi nu prezintă avantaje semnificative în ceea ce privește sporul de productivitate pe care-l poate aduce.

Aplicația 2. Pentru ca un tub fluorescent să fie corespunzător calității prescrise, durata lui de ardere trebuie să fie de 2 000 ore. În scopul verificării calității loturilor zilnice de tuburi fluorescente produse, se extrage un eșantion de 20 de produse. Întrucît abaterea medie pătratică ce caracterizează variația duratei de ardere a tuburilor în populația originară nu se cunoaște, se va utiliza abaterea medie pătratică de sondaj și se va folosi testul lui Student de forma (6.4).

Rezultatele determinării duratei de ardere a celor 20 tuburi sînt arătate în tabelul 6.2. folosindu-se momentul convențional de ordinul întîi ($A = 9$).

Tabelul 6.2

Durata de ardere (x_i)	Numărul tuburilor cu durata x de ardere (f_i)	$\xi_i = \frac{x_i - A}{i}$	$\xi_i f_i$	$\xi_i^2 f_i$
1960	2	−2	−4	8
1980	4	−1	−4	4
2000	9	0	0	0
2020	3	1	3	3
2040	2	2	4	8
Total	20	0	−8 +7 −1	23

Media și dispersia de sondaj au valorile

$$\bar{x} = \frac{\Sigma \xi_i f_i}{n} \cdot 20 + 2000 = \frac{-1}{20} \cdot 20 + 2000 = 1999$$

$$s^2 = 20^2 \left[\frac{\xi_i^2 f_i}{n} - \frac{1}{n} (\Sigma \xi f_i)^2 \right] = 20^2 \cdot \left[\frac{23}{20} - 0{,}05^2 \right] \approx 460;$$

deci, abaterea medie pătratică este:

$$s = \sqrt{460}$$

Ne punem întrebarea dacă există o diferență semnificativă între media de sondaj $\bar{x} = 1999$ și media teoretică $m = 2000$ ore.

Aplicăm criteriul lui Student în forma dată de expresia (6.4):

$$t = \frac{\bar{x} - m}{s} \sqrt{n-1} = \frac{1999 - 2\,000}{\sqrt{460}} \cdot \sqrt{20-1} = \frac{(-1)\sqrt{19}}{\sqrt{460}} = \frac{-4{,}36}{21{,}5} \simeq -0{,}20.$$

Probabilității $p = 0{,}97$ și numărului gradelor de libertate $n - 1 = 19$ îi corespunde, potrivit tabelelor speciale, $t = 2{,}0$.

Întrucît $|-0{,}20 < 2{,}0|$, se poate afirma, cu o probabilitate de 0,97, că durata de ardere medie a tuburilor fluorescente este corespunzătoare.

Scanned with OKEN Scanner

6.5 COMPARAREA VALORILOR MEDII DE SONDAJ ÎN CAZUL A DOUĂ EȘANTIOANE MICI

6.5.1 Ipoteza de egalitate a dispersiilor necunoscute

Să considerăm două eșantioane $x_1, x_2, \ldots, x_{n1}$ și $x'_1, x'_2, \ldots, x'_{n2}$ din care cel puțin unul este de volum redus.

După cum se știe, pentru aceste două eșantioane putem defini indicatorii respectivi de estimație a valorilor medii și a dispersiilor:

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n} x_i, \qquad \tilde{s}_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n}(x_i - \bar{x}_1)^2;$$
$$\bar{x}_2 = \frac{1}{n_2}\sum x'_i, \qquad \tilde{s}_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n}(x'_i - \bar{x}_2)^2. \tag{6.5}$$

Ne punem problema să aflăm dacă cele două eșantioane provin din una și aceeași populație. Aceasta înseamnă a verifica ipoteza diferenței nule, constînd în presupunerea că mediile m_1 și m_2 ale populațiilor originare sînt egale $(m_1 - m_2 = 0)$.

Verificarea ipotezei diferenței nule în acest caz se face cu ajutorul testului lui Student, în formă modificată:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_0\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{6.6}$$

în care indicatorul s_0 reprezintă abaterea medie pătratică a celor două eșantioane considerate laolaltă:

$$s_0 = \sqrt{\frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x'_1 - \bar{x}_2)^2}. \tag{6.7}$$

Relația (6.7) o putem scrie și sub o altă formă, comodă pentru executarea calculelor mecanizat și pe care o utilizăm mai jos, la calculele din aplicația 3.

$$s_0 = \sqrt{\frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1} x_i^2 - \frac{\left(\sum_{i=1}^{n_1} x_i\right)^2}{n_1} + \sum_{i=1}^{n_2} x_2' - \frac{\left(\sum_{i=1}^{n_2} x'_i\right)^2}{n_2}\right]}. \tag{6.7'}$$

Pe baza rezultatelor teoretice care ne sînt acum cunoscute putem afirma că dacă cele două eșantioane provin din aceeași populație orginiară, atunci s_0 va fi o estimație a lui σ.

Indicatorul de estimație s_0 are un număr $\nu = n_1 + n_2 - 2$ de grade de libertate, întrucît atît $\bar{x}_1$ cît și $\bar{x}_2$ sînt calculați cu ajutorul datelor de observație.

Scanned with OKEN Scanner

Aplicația 3. Este supusă observării și se înregistrează productivitatea orară (w) a doi muncitori care lucrează la două strunguri de tip diferit. Se fac șase înregistrări ale productivității muncii la primul muncitor și, respectiv, patru la al doilea. Datele obținute sînt cele din tabelul 6.3.

Tabelul 6.3

NIVELUL PRODUCTIVITĂȚII ORARE A MUNCII — ÎN TIMP DE MUNCĂ — A DOI MUNCITORI ÎN ȘASE ȘI RESPECTIV PATRU ÎNREGISTRĂRI

w_{1i}	w_{2i}	w_{1i}^2	w_{2i}^2
4	3	16	9
5	5	25	25
4,5	4,5	20,25	20,25
6	6	36	36
5,5	—	30,25	—
4	—	16	—

Indicatorii necesari pentru calculul abaterii medii pătratice întrunite pe cele două eșantioane, s_0, sînt următorii:

Eșantionul 1

$$n_1 = 6;\ \bar{w}_1 = \frac{29}{6} = 4{,}83;$$

$$\frac{\left(\sum_{i=1}^{n_1} w_{1i}\right)^2}{n_1} = \frac{29^2}{6} = \frac{841}{6}$$

Eșantionul 2

$$n_2 = 4;\ \bar{w}_2 = \frac{18{,}5}{4} = 4{,}625$$

$$\frac{\left(\sum_{i=1}^{n_2} w_{2i}\right)^2}{n_2} = \frac{342{,}25}{4}$$

Prin urmare avem:

$$s_0 = \sqrt{\frac{1}{n_1 + n_2 - 2}\left[143{,}50 - \frac{841}{6} + 90{,}25 - \frac{342{,}25}{4}\right]} = \sqrt{\frac{8{,}0175}{8}} = 1{,}01.$$

Acum putem calcula și valoarea lui t definită de relația (6.6) și s-o comparăm cu valoarea teoretică tabelată, în cazul numărului gradelor de libertate $\sigma = n_1 + n_2 - 2$.

Avem:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_0\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{4{,}83 - 4{,}625}{1{,}01\sqrt{\frac{1}{6} + \frac{1}{4}}} = \frac{0{,}205}{0{,}65145} = 0{,}314.$$

Comparînd valoarea de sondaj obținută cu valoarea teoretică din tabele speciale, putem constata că ea este mai mică decît aceasta din urmă, în cazul unui nivel de semnificație de 2‰ și al numărului de grade de libertate egal cu 8. Devine astfel clară și concluzia pe care trebuie s-o tragem în cazul examinat: ipoteza diferenței nule pe care am formulat-o în exemplul considerat trebuie admisă, considerîndu-se astfel că diferența dintre cele două valori medii ale productivității muncii, calculate pe baza datelor de sondaj, este

Scanned with OKEN Scanner

cu totul întîmplătoare. În alţi termeni, se poate afirma că tipurile diferite de strunguri la care lucrează cei doi muncitori nu exercită influenţă semnificativă asupra nivelului productivităţii acestora.

Este important de precizat modul în care se poate construi *intervalul de încredere* pentru diferenţa dintre două medii de sondaj, eşantioanele fiind mici. Trebuie să precizăm că acestui scop îi poate servi repartiţia lui Student, aplicată în acelaşi mod în care am văzut că s-a aplicat repartiţia normală. Intervalul respectiv va fi cuprins între următoarele limite:

$$(\bar{x}_1 - \bar{x}_2) - t \cdot s_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ şi } \bar{x}_1 - \bar{x}_2 + t \cdot s_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \tag{6.8}$$

În aplicaţia pe care am considerat-o mai sus (vezi tabelul 6.3) avem diferenţa dintre medii $(\bar{x}_1 - \bar{x}_2)$ egală cu 0,205.

Întrucît la 8 grade de libertate, în cazul unui nivel de semnificaţie, să zicem de 5%, valoarea teoretică a lui t este egală cu 2,110, limitele intervalului de încredere de 95% pentru diferenţa dintre valorile medii vor fi:

$$0{,}205 - 2{,}110 \cdot 0{,}651 \text{ şi } 0{,}205 + 2{,}110 \cdot 0{,}651$$

sau

$$0{,}205 \pm 1{,}373.$$

După cum se vede, diferenţa dintre valorile medii va oscila cu o probabilitate de 0,95 între $-1{,}168$ şi $+1{,}578$ unităţi de măsură.

Se poate întîmpla deci ca în unele eşantioane productivitatea medie a primului muncitor să fie mai mică decît a celui de-al doilea, în altele însă ne putem aştepta ca situaţia să fie inversă.

După cum am văzut, aplicarea criteriului de semnificaţie (6.6) ne-a dus la concluzia că ipoteza diferenţei nule dintre cele două medii poate fi admisă şi în acest sens se poate observa că în intervalul de încredere pe care l-am construit mai sus este cuprinsă şi valoarea zero a diferenţei.

6.5.2 Ipoteza de inegalitate a dispersiilor necunoscute

Referindu-se din nou la exemplul arătat cu cei doi muncitori, ale căror productivităţi au fost comparate, noi am admis ipoteza egalităţii dispersiilor ce caracterizează populaţiile originare din care s-au prelevat cele două eşantioane. Se ştie însă că în cercetările practice o asemenea ipoteză trebuie, în prealabil, verificată.

R. Fisher a studiat această problemă şi a introdus un criteriu statistic cu ajutorul căruia poate fi verificată justeţea ipotezei de egalitate a dispersiilor.

Legea teoretică de repartiţie a unei variabile aleatoare de forma mărimii lui Fisher (3.76) ne este cunoscută, vezi (3.77).

În procesul cercetării şi generalizării informaţiilor oferite de eşantioane mici, mărimea de sondaj „F“ se compară cu mărimea teoretică, corespunzătoare unui nivel de semnificaţie dat şi unui număr de grade de libertate $\nu_1 = n_1 - 1$ al numărătorului şi $\nu_2 = n_2 - 1$ — al numitorului.

Scanned with OKEN Scanner

Dacă mărimea „F“ de sondaj este mai mare decît cea teoretică, atunci rezultatul sondajului trebuie interpretat ca fiind *semnificativ*, în sensul că diferența puternică de variabilitate în cele două eșantioane nu trebuie atribuită fluctuațiilor aleatoare de sondaj. Cercetătorul va trebui să conchidă că există cauze specifice fiecărei populații care implică și tipuri diferite de împrăștiere ale caracteristicii studiate.

Observație. De obicei, criteriul „F“ este folosit în analize care necesită doar aprecieri ale abaterilor într-un singur sens și de aceea, tabelele cu repartițiile teoretice ale lui „F“ sînt date tocmai în această formă. În practică se ivesc însă și cazuri în care ne pot interesa, în aceeași măsură, abaterile în ambele sensuri, pozitive și negative. În acest caz, dacă, de pildă, alegem un nivel de semnificație de 5%, vom căuta în tabele valoarea lui F corespunzătoare unui nivel de semnificație egal cu 2,5%.

Din cele arătate cu privire la testul lui Fisher am reținut ideea principală: dacă dispersiile necunoscute în cele două populații originare sînt aceleași — testul nefiind semnificativ — atunci este aplicabil și testul lui Student în condițiile arătate.

Problema care se pune însă imediat este aceea de a ști cum vom proceda dacă raportul F dintre dispersiile s_1^2 și s_2^2 este semnificativ.

Teoria sondajului bazată pe eșantioane mici ne oferă și în acest caz mai multe procedee de analiză.

Unul din ele — dealtfel și cel mai simplu — constă în a admite în calcule o anumită aproximare.

Îl vom descrie după ce în prealabil vom reaminti o expresie de care ne-am mai ocupat:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

care, după cum știm, definește o variabilă aleatoare cu repartiția normală, avînd media 0 și dispersia 1. În construcția ei remarcăm mediile de sondaj $\bar{x}_1$ și $\bar{x}_2$ a două eșantioane de date despre care știm că au repartiții normale, dispersiile acestor medii de sondaj fiind respectiv $\frac{s_1^2}{n_1}$ și $\frac{s_2^2}{n_2}$.

Procedeul de aproximare se bazează pe ideea că variabila (d) are, aproximativ, aceeași repartiție ca cea a lui Student, cu un număr de grade de libertate determinat în felul următor:

$$\nu = \frac{1}{\frac{u^2}{n_1 - 1} + \frac{(1-u)^2}{n_2 - 1}} \tag{6.9}$$

în care, u este dat de expresia:

$$u = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}. \tag{6.10}$$

Calculele care ne conduc la determinarea numărului gradelor de libertate sînt simple. Obținînd, pe baza lor, valoarea variabilei d, considerată ca avînd

Scanned with OKEN Scanner

repartiția Student cu ν grade de libertate, putem construi, în continuare, intervalul de încredere pentru diferența dintre mediile m_1 și m_2 ale celor două populații originare, adică:

$$(\bar{x}_1 - \bar{x}_2) - d \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < m_1 - m_2 < (\bar{x}_1 - \bar{x}_2) + d \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \quad (6.11)$$

Pentru a ilustra calculul intervalului de încredere, în ipoteza unor dispersii inegale, vom recurge la o altă aplicație în care se modifică, de fapt, datele aplicației precedente.

Aplicația 4. Să presupunem că cele patru informații privind productivitatea orară a muncii celui de-al doilea muncitor — vezi tabelul 6.3 — sînt altele, seria fiind următoarea: 1; 6; 3,5; 8. Se cere construirea intervalului de încredere în care, cu o probabilitate de 0,95, se va localiza diferența dintre mediile m_1 și m_2.

În cazul expus în această aplicație, deși media măsurătorilor privind productivitatea muncii celui de-al doilea muncitor este aceeași ca în prima aplicație ($w_2 = 4{,}625$), dispersia are totuși o valoare mult mai ridicată, și anume $s_2^2 = 9{,}23$.

Într-adevăr,

$$\tilde{s}_2^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n_2} w_{2i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} w_{2i}\right)^2\right] = \frac{1}{3}\left[113{,}25 - \frac{342{,}25}{4}\right] = \frac{27{,}6875}{3} = 9{,}23.$$

Întrucît dispersia productivității muncii primului muncitor este egală cu 0,666, rezultă că valoarea raportului $F = \frac{9{,}23}{0{,}666} \simeq 13{,}8$ este, statistic, *net semnificativă*. Această concluzie rezultă din faptul că valoarea de sondaj a acestui raport depășește cu mult chiar și valoarea teoretică ($F_{0,99} = 12{,}06$) corespunzătoare unui nivel de semnificație de 1% și datelor din aplicația 4 privind numărul gradelor de libertate, adică 3 și respectiv 5.

Prin urmare, sîntem în situația de a aplica inegalitatea (6.11) la construirea intervalului de încredere. În acest scop vom determina, mai întîi, numărul gradelor de libertate ale variabilei aleatoare d, considerată drept o variabilă care urmează repartiția Student, folosind relațiile (6.10) și (6.11).

În cazul datelor noastre avem:

$$u = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2 n_2} = \frac{0{,}666/6}{0{,}666/6 + 9{,}23/4} = \frac{0{,}111}{0{,}111 + 2{,}3075} = 0{,}045,$$

iar numărul gradelor de libertate este:

$$\nu = \frac{1}{u^2/n_1 - 1 + (1-u)^2/n_2 - 1} = \frac{1}{0{,}045^2/_{6-1} + (1-0{,}045)/_{4-1}} = \frac{1}{0{,}3044} \simeq 3.$$

Consultînd tabelele speciale privind repartițiile Student, constatăm că unui nivel de semnificație de 5% și unui număr de trei grade de libertate îi corespunde o valoare teoretică a lui d egală cu 3,182.

Scanned with OKEN Scanner

Putem acum construi intervalul de încredere cerut, pe baza datelor din aplicația 4:

$$(\bar{x}_1 - \bar{x}_2) - 3{,}182\sqrt{\frac{0{,}666}{6} + \frac{9{,}23}{4}} < m_1 - m_2 < (x_1 - x_2) +$$

$$+ 3{,}182\sqrt{\frac{0{,}666}{6} + \frac{9{,}23}{4}}.$$

Un calcul elementar ne arată că diferența dintre mediile populațiilor originare din care au fost prelevate cele două eșantioane se localizează, cu o probabilitate de 0,95, în cadrul următoarelor limite:

$0{,}205 - 3{,}182 \cdot 1{,}555$ și $0{,}205 + 3{,}182 \cdot 1{,}555$

sau:

$0{,}205 \pm 4{,}948$.

După cum se vede, limitele intervalului sînt mari, în interiorul lor găsindu-se și valoarea zero, ceea ce confirmă din nou, probabilistic, ipoteza diferenței nule dintre cele două medii ale populațiilor originare.

Aplicația 5[2]. Pentru determinarea conținutului în proteină al recoltei de grîu din regiunile de vest ale Canadei s-a recurs la o cercetare pe bază de două eșantioane mici. Primul eșantion, format din 8 valori ($n_1 = 8$), caracterizează conținutul în proteină al grîului recoltat la ferma Saskatchewan, iar al doilea, format din 14 valori ($n_2 = 14$), caracterizează același conținut însă pentru grîul recoltat la ferma Manitoba.

Se cere aplicarea testului t al lui Student și a testului Fisher.

Datele privind valorile constatate în cele două eșantioane, ca și indicatorii pentru calculul testelor amintite sînt:

Eșantionul 1

$n_1 = 8$

$x_i = 15{,}1;\ 14{,}3;\ 11{,}5;\ 14{,}5;\ 15{,}4;\ 12{,}5;\ 14{,}6;\ 16{,}6.$

$\sum_{i=1}^{n_1} x_i = 114{,}5;\quad \bar{x}_1 = 14{,}31$

$\sum_{i=1}^{n_1} x_i^2 = 1657{,}13$

$\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 = 1657{,}13 - \frac{(114{,}5)^2}{8} = 18{,}35$

$\tilde{s}_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (x_i - \bar{x}_2)^2 = \frac{18{,}35}{7} = 2{,}6214$

$\tilde{s}_{\bar{x}_1}^2 = \frac{2{,}6214}{8} = 0{,}328$

Eșantionul 2

$n_2 = 14$

$x_i' = 12{,}2;\ 12{,}5;\ 11{,}2;\ 12{,}6;\ 11{,}0;\ 11{,}6;\ 12{,}0\ 12{,}5;\ 11{,}8;\ 12{,}4.$

$\sum_{i=1}^{n_2} x_i' = 167{,}6;\quad \bar{x}_2 = 11{,}97$

$\sum_{i=1}^{n_2} x_i'^2 = 2010{,}20$

$\sum_{i=1}^{n_2} (x_i' - \bar{x}_2)^2 = 2010{,}20 - \frac{(167{,}6)^2}{14} = 3{,}79$

$\tilde{s}_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} (x_i' - \bar{x}_2)^2 = \frac{3{,}79}{13} = 0{,}2915$

$\tilde{s}_{\bar{x}_2}^2 = \frac{0{,}2915}{14} = 0{,}021.$

[2] Această aplicație este luată din: CYRIL H. GOULDEN, *Methods of Statistical Analysis*, ed. a II-a, New York, London, 1956, p. 58.

Scanned with OKEN Scanner

Pentru a testa dacă dispersiile în cele două eşantioane diferă semnificativ calculăm mărimea lui Fisher:

$$F = \frac{s_1^2}{s_2^2} = \frac{2{,}6214^1}{0{,}2915} = 9{,}0.$$

Din tabele speciale privind repartiţiile teoretice ale mărimii lui Fisher se constată, de pildă, că unui nivel de semnificaţie de 5% şi numerelor de grade de libertate egale cu 7 şi respectiv 13, le corespunde o mărime a raportului F egală cu 2,84, iar unui nivel de semnificaţie de 1% şi aceloraşi grade de libertate le corespunde $F = 4{,}44$. Deci, la ambele nivele de semnificaţie, mărimile teoretice fiind mai mici decît mărimea de sondaj, putem conchide că dispersiile în cele două populaţii originare sînt puternic diferite.

Cît priveşte diferenţa dintre cele două medii de sondaj $\bar{x}_1 - \bar{x}_2 = 14{,}31 - -11{,}97 = 2{,}34$ se constată, prin aplicarea testului lui Student, că şi ea este semnificativă, ipoteza diferenţei nule dintre ele urmînd a fi astfel respinsă. Lăsăm pe seama cititorului — cu titlu de exerciţiu — efectuarea calculelor respective.

6.6. CÎTEVA OBSERVAŢII GENERALE PE MARGINEA APLICAŢIILOR REFERITOARE LA EŞANTIOANE MICI

În aplicaţiile considerate mai înainte, volumele eşantioanelor au oscilat între 4 şi 14 cazuri observate. Deşi foarte mici, ele ne-au permis totuşi să facem unele estimaţii ale valorilor tipice ce caracterizează populaţiile originare şi aceasta pe baza testelor de semnificaţie arătate.

Se impun, de aceea, cîteva observaţii pe marginea aplicaţiilor referitoare la sondaje de volum redus.

1°. În legătură cu oportunitatea şi legitimitatea aplicării acestor teste, cititorul îşi poate pune, în primul rînd, problema de a şti care este limita superioară pînă la care volumul eşantionului mai poate fi considerat ca fiind mic.

Asupra acestei probleme noi am anticipat cu alt prilej, subliniind că din practica cercetărilor pe bază de sondaj a rezultat şi rezolvarea ei, considerîndu-se astfel că un eşantion poate fi apreciat ca fiind mic, dacă numărul cazurilor observate este egal sau mai mic de 30.

Putem preciza acum că această apreciere — oarecum empirică — are o justificare teoretică. Într-adevăr, cercetînd tabelele în care sînt calculate repartiţiile teoretice ale testului Student, la diferite nivele de semnificaţie şi grade de libertate, se poate observa că valorile lui t, corespunzătoare unui număr de grade de libertate mai mare de 30 pînă la $+\infty$, nu diferă semnificativ de valorile ce corespund eşantioanelor cu maximum 30 grade de libertate.

2°. Din observaţia de mai sus rezultă, implicit, concluzia că eşantioanele care înglobează mai mult de 30 măsurători pot fi studiate şi aproximate pe baza repartiţiei normale, întrucît repartiţia lui Student se apropie mult de o repartiţie normală în cazul în care numărul cazurilor observate creşte necontenit.

3° Este deosebit de important să subliniem din nou că, în general, aplicînd testul lui Student noi facem ipoteza că populaţia originară se supune unei repartiţii normale. În legătură cu această problemă este de reţinut următorul

Scanned with OKEN Scanner

aspect important ce a rezultat mai ales din cercetările practice pe bază de sondaje mici. În cazul în care populaţia originară se abate de la normalitate, criteriul lui Student devine deosebit de sensibil. Se impune, de aceea, mult discernămînt în aplicarea lui, mai ales în cazul în care cercetătorul are suficiente temeiuri să considere că repartiţia populaţiei originare este o repartiţie specifică.

6.7 CÎTEVA PRECIZĂRI ŞI CONCLUZII FINALE PRIVIND SONDAJUL STATISTIC

În cele două capitole rezervate teoriei şi aplicaţiilor sondajului statistic n-am putut epuiza, desigur, numeroasele probleme teoretico-metodologice ca şi cele practice pe care acest sondaj le implică. Totuşi, socotim că au fost abordate, într-o viziune aplicativă, cele mai importante teoreme de teoria probabilităţilor şi statistică matematică care fundamentează şi justifică efectuarea cercetărilor — pe bază de sondaje — asupra diferitelor populaţii statistice unidimensionale aşa cum le întîlnim în practica statistică obişnuită. Scheme de sondaj mai complexe, referitoare la populaţii *multidimensionale*, cititorul interesat le poate găsi în numeroase studii şi lucrări de specialitate. [3]

De asemenea, în lucrarea [90], consacrată problemelor pe care le ridică studiul corelaţiei, pot fi găsite expuneri privind probleme de estimare şi testare a principalelor valori tipice ale populaţiilor originare pe baza repartiţiilor de sondaj multidimensionale, considerîndu-se în context şi erorile standard respective.

În încheierea acestui capitol vom face cîteva precizări şi formula unele concluzii finale, de ansamblu, privind sondajul statistic.

6.7.1 **Problema estimării**

Sintetizînd metodologia expusă pînă acum privind estimările parametrilor ce caracterizează populaţiile originare putem reţine că, în general, indiferent de volumul eşantionului, principalul instrument de investigare folosit în procesul estimării l-a constituit construcţia *momentelor de sondaj*, ponderile implicate în context fiind, în locul probabilităţilor, *frecvenţele empirice*, de sondaj.

Vom preciza acum că există o metodă mai generală ce poate fi folosită la estimarea diferiţilor parametri ai populaţiilor originare — *metoda verosimilităţii maxime*.

De regulă, aplicarea acestei metode conduce, în final, la formularea unor indicatori statistici de înaltă calitate, bucurîndu-se de o serie de proprietăţi deosebite, printre care, în primul rînd, aceea de a oferi, după caz, estimaţii *consistente* sau foarte uşor — nesemnificativ — *deplasate*, şi totodată, *eficiente*

[3] Vezi, de pildă [60, p. 353 şi 359] şi [85 p. 540–554 şi 570–572].

Scanned with OKEN Scanner

și *suficiente*. Metoda verosimilității maxime introduce deci o serie de raționamente ce vizează aprecierea estimațiilor obținute pe baza acestor patru criterii.

Expunem, succint, esența acestei metode ca și a celor patru criterii amintite.

Fie $x_1, x_2, \ldots, x_n$ un eșantion aleator prelevat dintr-o populație a cărei densitate de repartiție este $f(x)$.

Probabilitatea eventimentului, constînd în faptul că fiecare valoare de sondaj se va plasa în intervalul $x_i - \frac{\varepsilon}{2}$ și $x_i + \frac{\varepsilon}{2}$ va fi egală, aproximativ, cu $f(x_i) \cdot \varepsilon$. Dar, în acest caz, probabilitatea ca prima valoare de sondaj să cadă în primul interval, a doua — în al doilea ș.a.m.d. — potrivit teoremei de înmulțire a probabilităților — va fi egală cu:

$$P = f(x_1) \cdot f(x_2) \ldots f(x_n) \cdot \varepsilon^n.$$

De aici deducem:

$$\frac{P}{\varepsilon^n} = f(x_1) \cdot f(x_2) \ldots f(x_n).$$

Logaritmînd această expresie obținem:

$$\ln \left(\frac{P}{\varepsilon^n} \right) = \ln \left[f(x_1) \cdot f(x_2) \ldots f(x_n) \right] = \sum_{i=1}^{n} \ln \left[f(x_i) \right] = L. \qquad (6.12)$$

Această expresie poartă denumirea de *funcție de verosimilitate*.

Principiul verosimilității maxime constă în aceea că, la alegerea și formularea unei ipoteze oarecare, cercetătorul se oprește asupra aceleia care face ca funcția (6.12) să ia valoarea maximă.

Determinarea maximului funcției de verosimilitate (6.12) se face potrivit regulilor obișnuite de cercetare a funcțiilor de cîteva variabile. De pildă, pentru valorile tipice cum sînt media și dispersia, avem:

$$\frac{\partial L}{\partial m} = 0 \text{ și } \frac{\partial L}{\partial \sigma^2} = 0.$$

Rezolvarea acestor ecuații implică aplicarea nemijlocită a metodei verosimilității maxime.

Să considerăm, de pildă, cazul unui eșantion oarecare extras dintr-o populație a cărei repartiție este normală, iar parametrii principali — media (m) și dispersia (σ^2) — sînt necunoscuți.

Să construim estimații de maximă verosimilitate pentru cei doi parametri necunoscuți ai populației originare.

Obținem:

$$L = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x_1-m)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_2-m)^2}{2\sigma^2}} \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-m)^2}{2\sigma^2}} =$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2};$$

$$ln\ L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (x_i - m)^2.$$

Scanned with OKEN Scanner

Diferențiind ln L, după m și σ^2, și egalînd cu zero derivatele parțiale obținute ajungem la următorul sistem de ecuații:

$$\begin{cases} \dfrac{\partial \ln L}{\partial m} = \dfrac{1}{\sigma^2} \sum\limits_{i=1}^{n} (x_i - m)^2 = 0 \\ \dfrac{\partial \ln L}{\partial (\sigma^2)} = -\dfrac{n}{2} \cdot \dfrac{1}{\sigma^2} + \dfrac{2}{2(\sigma^2)^2} \cdot \sum\limits_{i=1}^{n} (x_i - m)^2 = 0. \end{cases}$$

Soluția sistemului ne oferă următoarele valori:

a) $m = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$, adică valoarea medie de sondaj, $(\bar{x})$;

b) $\sigma^2 = \dfrac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n}$, adică dispersia de sondaj, s^2.

După cum se vede, aplicarea metodei verosimilității maxime ne conduce la concluzia că media de sondaj și dispersia de sondaj sînt indicatori de estimație pentru parametrii respectivi ai populației originare.

Desigur ne punem întotodeauna și problema de a ști care este natura acestor estimații sub aspectul celor patru criterii: al consistenței, nedeplasării, eficienței și suficienței.

O estimație se spune că este *consistentă* dacă indicatorul respectiv, obținut pe baza informației de sondaj, converge „în probabilitate" — în virtutea legii numerelor mari — către valoarea sa teoretică, pe măsură ce numărul observațiilor crește necontenit. Expresia „în probabilitate" este echivalentă cu expresia „frecvența relativă a evenimentului tinde către probabilitate, cînd numărul cazurilor observate crește la infinit". De pildă, dacă probabilitatea inegalității $\left| \dfrac{f}{n} - p \right| < \varepsilon$ se apropie de 1, cînd $n \to \infty$, mărimea ε fiind oricît de mică, putem afirma că $\dfrac{f}{n}$ converge „în probabilitate" către p.

Se consideră, astfel, ca fiind estimații consistente pentru parametrii respectivi ai populației originare, următorii indicatori de sondaj: media și dispersia de sondaj: proporția de sondaj a unei caracteristici-atribut; coeficienții de asimetrie și de exces ș.a.

În cazul în care valorile tipice de sondaj sînt concordante cu valorile corespunzătoare ale populației originare, ele constituie estimații *nedeplasate* ale acestora. De pildă, valoarea medie a mediei de sondaj este egală cu valoarea medie a populației originare $M(\bar{x}) = m$; ea constituie deci o estimație nedeplasată. Nu același lucru se poate spune despre valoarea medie a dispersiei de sondaj. Se știe, astfel, că $M(s^2) \neq \sigma^2$. Pentru a înlătura eroarea sistematică și a obține o estimație *nedeplasată*, dispersia de sondaj trebuie înmulțită cu $\dfrac{n}{n-1}$:

$$\tilde{s}^2 = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

Scanned with OKEN Scanner

O estimaţie nedeplasată, care are cea mai mică dispersie în jurul parametrului estimat al populaţiei originare, se spune că este *eficientă*. Pentru aplicarea justă a criteriului eficienţei, rolul hotărîtor îl are caracterul legii de repartiţie a populaţiei originare.

O estimaţie eficientă pentru o anumită lege de repartiţie nu este identică cu o estimaţie eficientă pentru o altă lege de repartiţie. Estimaţia eficientă este totodată şi o estimaţie consistentă.

În sfîrşit, se spune că o estimaţie este *suficientă* (exhaustivă), atunci cînd asigură folosirea deplină a informaţiei conţinute în eşantion. Nici o altă estimaţie — cu excepţia celei suficiente — nu poate oferi informaţii suplimentare cu privire la parametrii estimaţi.

6.7.2 Conceperea planului de sondaj şi problema omogenităţii populaţiei originare

Orice sondaj statistic trebuie să se efectueze după un plan bine organizat, conceput — în cazul cînd sînt luate în studiu fenomene ale vieţii social-economice — de către statisticieni economişti, specializaţi în problemele teoriei şi practicii sondajelor. Cunoştinţele de teoria probabilităţilor şi statistică matematică sînt, evident, necesare dar nu şi suficiente. Un plan de sondaj trebuie ales astfel încît să conducă la rezultate cît mai precise şi, totodată, la cheltuieli materiale, de timp şi băneşti, *minime*.

1°. De regulă, populaţia originară trebuie divizată în subpopulaţii sau grupe — straturi tipice, astfel constituite încît în interiorul fiecăreia caracteristica supusă cercetării să fie repartizată cît mai *omogen* cu putinţă, mai ales sub aspectul variabilităţii pe care o manifestă.

Aceasta este necesar întrucît adeseori în practică nu putem avea suficiente informaţii sau criterii pentru a emite anumite ipoteze cu privire la *gradul de omogenitate* al populaţiei originare luate în ansamblul ei. Ne putem deci afla ori în faţa unor populaţii originare omogene, ori a unora puternic eterogene.

Este sigur că dacă populaţia orginară este omogenă şi procedeul de sondaj este riguros aleator, nu prezintă nici un avantaj, din punct de vedere teoretic, formarea de subeşantioane pe anumite straturi de populaţie, întrucît în acest caz orice tehnică aleatoare simplă de sondaj oferă fiecărui element al populaţiei aceeaşi şansă de a fi ales în eşantion. Totuşi, în acest caz nu este exclusă necesitatea — impusă de anumite motive de interes practic — să alegem, la întîmplare, *grupe* de elemente, cu condiţia să avem suficiente motive de a presupune că elementele acestor grupe sînt repartizate, la întîmplare, în interiorul fiecăreia din ele.

Devine limpede că, prelevînd un eşantion aleator de asemenea *grupe*, fiecărui element al populaţiei i se asigură aceeaşi şansă de a fi ales, astfel încît putem afirma că ambele tehnici de sondaj sînt echivalente.

Practica a arătat însă că uneori însăşi natura populaţiilor originare este cea care face ca elementele ei să nu se repartizeze în grupe sau straturi omogene. De pildă, în cazul unor sondaje dintr-o populaţie umană avem apriori

Scanned with OKEN Scanner

motive serioase — şi aceasta în mod uzual — de a considera că o anumită caracteristică pe care o investigăm să difere, să zicem, de la aşezări urbane la cele rurale, astfel încît populaţia este evident eterogenă şi în limitele unui teritoriu dat există, în mod firesc, o corelaţie pozitivă între elementele învecinate ale eşantionului.

Desigur, în acest caz tehnicile noastre de sondaj vor trebui adaptate în mod corespunzător. Poate fi sugerată, de pildă, o tehnică de sondaj care să asigure prelevarea în eşantion doar a unui singur membru din fiecare familie. Procedînd astfel, putem exclude cazurile întîmplătoare de coincidenţe sau asemănări care sînt mai probabile în sînul unei singure familii decît între familii diferite. Subliniem, totuşi, că în cazul unui asemenea fenomen de neomogenitate a populaţiei originare tehnica propusă ar putea să ducă sau ar putea să nu ducă la reducerea erorii standard, totul depinzînd de caracterul real al eterogenităţii datelor originare.

După cum se vede, un rol hotărîtor îl pot avea pregătirea şi experienţa multilaterală a cercetătorului în a face anumite ipoteze — mai mult sau mai puţin plauzibile — cu privire la natura şi tipul populaţiilor originare.

În funcţie de aceste ipoteze, el va alege apoi, cu suficientă încredere, tehnicile de sondaj corespunzătoare.

Rezultă astfel că în practică se pot ivi serioase dificultăţi în a decide care anume tehnică de sondaj este preferabilă.

2°. În cazul în care populaţia este divizată în grupe tipice, la baza grupării trebuie să stea caracteristicile tipice, esenţiale, ale fenomenului studiat în conformitate cu sarcinile concrete ce revin cercetării pe bază de sondaj. Nu există reguli precise pentru determinarea grupelor, există însă o serie de principii generale de care trebuie să se ţină seama pentru a asigura grupării un caracter riguros ştiinţific. Aici intervine rolul cel mai important al celui ce elaborează planul de sondaj, care bazîndu-se pe principiile de mai sus, pe experienţa lui şi pe cunoaşterea profundă a domeniului în care se face cercetarea trebuie să facă cea mai justă grupare a populaţiei originare.

Odată grupele stabilite, trebuie să se fixeze numărul unităţilor care vor fi extrase din fiecare grupă, în raport cu gradul de omogenitate a grupei, ceea ce implică de asemenea o experienţă deosebită din partea celui care întocmeşte planul lucrării.

3°. După cum am văzut, numerele de extracţii din fiecare grupă depind de dispersia valorilor din fiecare grupă. Această dispersie este însă necunoscută. Este adevărat că pentru aplicarea sondajului nu avem nevoie de valoarea exactă a dispersiei, ci numai de o apreciere aproximativă a ei. Această apreciere nu o poate face însă decît un bun cunoscător al domeniului cercetat, pe baza unei anchete prealabile de mai mică întindere, ca şi a diferitelor studii de specialitate etc. Trebuie să subliniem că aceasta este dificulatatea principală în efectuarea cercetărilor pe bază de sondaj, depăşind cu mult dificultăţile legate de aplicarea tehnicii culegerii materialului statistic sau a formulelor matematice care variază foarte puţin de la un plan la altul şi pot fi uşor înlăturate prin aplicarea, aproape automată, a unor reguli fixe, existente în toate tratatele de specialitate.

Scanned with OKEN Scanner

CAPITOLUL 7
Ritm și tendințe

7.1 SERIILE CRONOLOGICE ÎN METODOLOGIA STATISTICĂ

7.1.1 Caracteristici generale

Metodele statistice de descriere și analiză a fenomenelor și proceselor din natură și societate, prezentate pînă acum, nu au luat în considerare modul de manifestare al acestora într-un anumit moment sau între anumite limite de timp. Repartițiile de frecvențe, de pildă, ca și dependența au fost privite dintr-un unghi de vedere care accepta o anumită stabilitate în structura repartițiilor, o permanență în interdependența dintre elementele constitutive ale sistemului cercetat.

Seriile cronologice, denumite și serii dinamice sau în timp, consitituie, în tipologia seriilor statistice, forma de prezentare ordonată a datelor în care se reflectă nivelul de manifestare a fenomenelor într-un anumit moment sau într-o anumită perioadă de timp.

Printr-un șir de indicatori numerici, seriile cronologice oglindesc procesul de schimbare și dezvoltare a unei anumite colectivități statistice, la intervale de timp succesive [85].

Spre deosebire de seriile de frecvență, a căror teorie are la bază postulatul stabilității, seriile cronologice oglindesc, printr-un șir de date, *evoluția* unui fenomen colectiv; în legătură cu structura lor nu se poate desprinde însă nici un element de permanență. „Orice desconsiderare a diferenței fundamentale dintre structura repartițiilor de frecvențe și aceea a seriilor în timp va constitui originea principală a oricăruia din așa-numitele eșecuri ale metodei statistice" [47].

Observația curentă ne încredințează că atît fenomenele și procesele naturale, cît mai ales cele cu caracter social-economic, variază în timp, desfășurîndu-se prin treceri dintr-o stare în alta. Mișcarea (schimbarea) fenomenelor în curgerea timpului este un proces dialectic ireversibil, care caracterizează în sens legiform realitatea. Schimbarea poate fi cantitativă sau calitativă: o creștere (acumulări cantitative în cadrul aceleiași determinări calitative), o transformare (schimbare calitativă), o dezvoltare (mișcare ascendentă, transformare calitativă în sens superior). În procesul mișcării se pot produce și transformări în sens regresiv; realitatea consemnează, de asemenea, acumulări cantitative și transformări calitative. Fiecare știință cercetează procesul acestor schimbări potrivit obiectului investigației sale specifice, dar ținînd seamă de legile dialecticii materialiste.

Scanned with OKEN Scanner

În seriile cronologice expresia numerică a mărimii poate fi o valoare absolută, relativă sau medie, de exemplu, evoluția unor indicatori economico-sociali (tabelul 7.1). Dinamica lor are caracter ascensional la unii, regresiv la alții, dar aceasta poate fi observată mai intuitiv într-o reprezentare grafică (fig. 7.1.).

Tabelul 7.1

DINAMICA UNOR INDICATORI AI DEZVOLTĂRII SOCIAL-ECONOMICE A ROMÂNIEI ÎN PERIOADA 1950—1974

Indicatori	U.m.	1950	1955	1960	1965	1970	1974
Populația ocupată	mii. pers.	8377,2	9382,2	9537,7	9684,0	9875,0	10070,1
Populația ocupată în agricultură	„	6208,7	6519,1	6233,1	5476,5	4848,6	4012,1
Produsul social	%	100	186	263	414	629	947
Producția globală industrială	„	100	202	340	649	1100	1900
Suprafața arabilă/1 tractor fizic	ha	684	420	222	121	91	83
Producția medie de grîu și secară la ha	chint.	8,0	10,2	12,1	19,7	14,4	20,8

După *Anuarul statistic al Republicii Socialiste România 1967* și *1975*.

Este suficient să urmărim evoluția acestor cîțiva indicatori pentru a desprinde ritmul dezvoltării, schimbările ce au loc în procesul făuririi societății socialiste multilateral dezvoltate — ca urmare a industrializării, mecanizării agriculturii, sporirii producției sociale, creșterii nivelului de trai —, trecerea de la acumulările cantitative la saltul calitativ.

Seriile cronologice au înfățișarea unui șir paralel de valori caracteristice de timp (x), cu rol de variabilă independentă și caracteristică a fenomenului (procesului) cercetat (y), care joacă rolul funcției. Vom avea deci relația $y = f(x)$, dar, spre deosebire de esența noțiunii de corelație, aici nu vorbim de o dependență cauzală, în care schimbarea argumentului funcției să determine schimbarea acesteia.

Pentru a înțelege mai bine aceasta trebuie să reținem, ca o caracteristică a seriilor cronologice, limitarea posibilității de a le construi numai la colectivități statistice omogene. „Aceasta înseamnă că se poate vorbi de schimbarea și dezvoltarea numerică — deci cantitativă — a unei *colectivități date*, atîta timp cît *esența ei calitativă rămîne aceeași*“ [85]. Dar aceasta nu înseamnă că o serie cronologică este neștiințific construită pentru perioade calitativ deosebite. O dovadă o constituie seriile cuprinse în tabelul 7.1, în care este oglindită evoluția unor indicatori specifici tocmai perioadei cu caracteristici calitativ deosebite în dezvoltarea societății noastre. Asigurarea unei colectivități statistice omogene pentru aplicarea metodei de analiză descriptivă a seriilor dinamice reclamă înlăturarea viciilor care ar putea afecta comparabilitatea indicatorilor constitutivi ai variantelor seriei, în ce privește volumul colectivității studiate, definirea unității de observare și alte condiții pe care le-am examinat atunci cînd am caracterizat etapa de observare statistică.

Scanned with OKEN Scanner

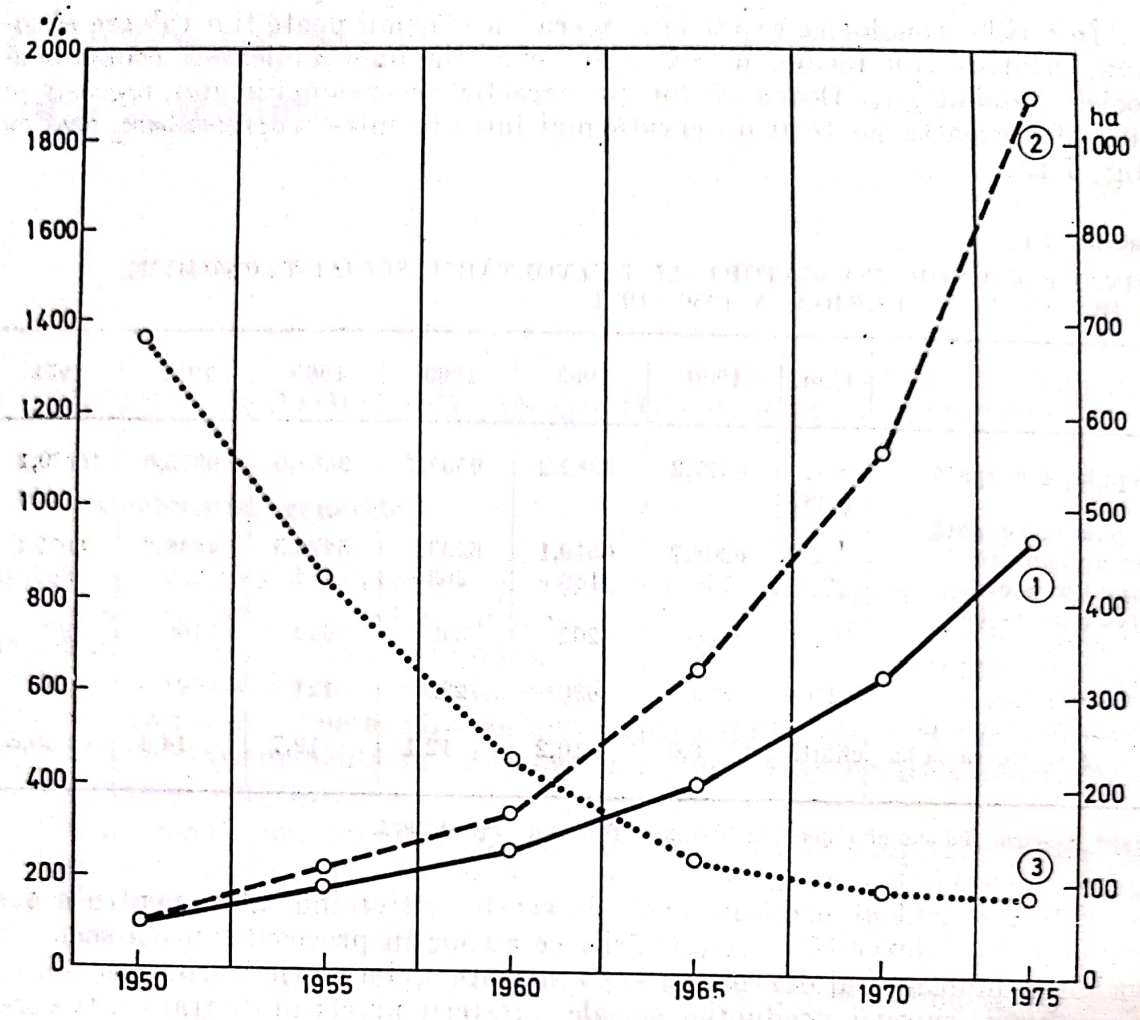


Fig. 7.1 Dinamica unor indicatori ai dezvoltării economice a României, în perioada 1950—1974. 1. Produsul social (%). 2. Producţie globală industrială (%) 3. Suprafaţa arabilă pe tractor fizic (ha). Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1975*.

O caracteristică a unei serii cronologice omogene rezultă şi din faptul că nivelul de dezvoltare a unui fenomen social-economic, într-o anumită perioadă, nu poate fi considerat independent, comparativ cu nivelul unei perioade anterioare, cîtă vreme *esenţa* fenomenului nu se schimbă în acel timp. Dimpotrivă, la aruncarea succesivă a unei monede, rezultatele unei experienţe nu influenţează rezultatele experienţei următoare. Întrucît schemele teoretice ale calculului probabilităţilor nu sînt în general aplicabile seriilor dinamice, statistica matematică pune la dispoziţia cercetării dezvoltării în timp a fenomenelor teoria proceselor stohastice, atunci cînd manifestarea fenomenului are un caracter întîmplător. Dacă se notează fenomenul respectiv cu x, apariţia valorilor posibile ale caracteristicii poate fi marcată, în diverse cazuri, printr-o probabilitate obiectiv condiţionată, de tipul t. De pildă, în statistica populaţiei şi a nivelului de trai, pentru diferite fenomene ca mortalitatea pe grupe de vîrstă, mărimea venitului într-o familie ş.a. se dă pentru fiecare perioadă t o anumită repartiţie a mărimilor respective ale caracteristicii. Cînd timpul este o succesiune (0,1,2 ...), procesul este cunoscut în general sub denumirea de lanţ. Procesele de tip Markov (lanţurile markoviene) şi procesele

Scanned with OKEN Scanner

staţionare constituie forme ale proceselor stohastice şi procedeele de tratare a acestor procese sînt folosite pe scară tot mai largă în cercetarea statistică a fenomenelor social-economice.

7.1.2 Clasificarea şi reprezentarea grafică a seriilor cronologice

Considerînd semnificaţiile ce le pot avea datele cuprinse într-o serie cronologică, se disting seriile de intervale şi seriile de momente.

În seriile de intervale, numite de unii serii diferenţiale, se reflectă evoluţia unui fenomen (proces) pe anumite perioade de timp ale manifestării sale, astfel cum a fost exemplificat cu indicatorii din tabelul 7.1. Nivelul lor, care reflectă rezultatele dezvoltării în unele domenii social-economice, a fost marcat pentru anumiţi ani; desigur că aceasta se poate face pe zile, luni sau alte fracţiuni temporale. Caracteristică pentru seriile de intervale este posibilitatea însumării mărimii succesive a indicatorilor; de exemplu, prin însumarea produsului social din lunile anului, obţinem produsul social al anului respectiv.

Seriile de momente, numite şi integrale, caracterizează nivelul de dezvoltare a fenomenelor la anumite momente de timp: numărul populaţiei (total, pe sexe, pe medii etc.) la 1 iulie a fiecărui an, numărul întreprinderilor industriale sau parcul de tractoare (total, pe judeţe, în unităţi fizice sau convenţionale) la sfîrşitul unui an (tabelul 7.2). Însumarea mărimii indicatorilor seriilor de momente este lipsită de sens, întrucît aceeaşi mărime ar fi luată în calcul de mai multe ori.

Între indicatorii din tabelul 7.2, numărul personalului este înregistrat ca nivel anual mediu. Asemenea exprimări valorice sînt, de pildă, şi producţia medie pe un lucrător sau muncitor, consumul anual mediu de energie electrică pe un muncitor industrial etc. Desigur că nici aceste valori medii anuale nu au sens să fie însumate.

Tabelul 7.2

EVOLUŢIA UNOR INDICATORI AI DEZVOLTĂRII SOCIAL-ECONOMICE A ROMÂNIEI, ÎN ANII 1970—1974

Indicatori	U.m.	1970	1971	1972	1073	1974
Populaţia totală (la 1 iulie)	mii loc.	20 253	20 470	20 663	20 828	21 029
Numărul mediu al personalului din industrie	mii pers.	2 066	2 202	2 324	2 485	2 660
Parcul total de tractoare, în unităţi fizice (la sfîrşitul anului)	mii buc.	107,3	114,2	115,6	116,5	116,8

După *Anuarul statistic al R. S. România 1975.*

Seriile de intervale sînt numite de unii autori şi cumulative, datorită caracteristicii menţionate mai înainte a însumării mărimilor cu care este înregistrat fenomenul în diferitele intervale de timp ale seriei. Alţi autori consideră seriile cumulative drept o clasă separată.

Scanned with OKEN Scanner

Tabelul 7.3 prezintă concomitent seria simplă de date ale unui fenomen și seria cumulată corespunzătoare, termenii uneia putîndu-se obține din termenii celeilalte prin însumare, respectiv scădere, cu excepția primului termen care este totdeauna comun ambelor serii.

Tabelul 7.3

DINAMICA VÎNZĂRILOR DE MĂRFURI CU AMĂNUNTUL PRIN COMERȚUL SOCIALIST ÎN ANII CINCINALULUI 1971—1975
(milioane lei—prețurile fiecărui an)

Ani (perioade)	Seria anuală	Seria cumulată
1971	105 940	105 940
1972 (1971—1972)	112 438	218 378
1973 (1971—1973)	121 290	339 668
1974 (1971—1974)	134 080	473 748
1975 (1971—1975)	143 419	617 167

După *Anuarul statistic al R. S. România 1975.*

Posibilitatea cumulării datelor prezintă importanță atît pentru alegerea intervalului de grupare, cît și pentru analiza statistico-economică. Prezentarea seriei din tabelul 7.3 scoate în evidență dezvoltarea fenomenului de la un an la altul și pe întregul cincinal. Adîncind analiza seriei, putem cunoaște caracterul ritmului de dezvoltare, factorii determinanți (creșterea producției, a consumului individual), rezultate de ordin social (creșterea nivelului de trai) etc.

Aspectul dinamic al fenomenelor urmărite în lucrările statistice solicită, pe o scară foarte largă, utilizarea reprezentărilor grafice, mai întîi pentru că ele ne ajută să sesizăm mult mai ușor și intuitiv sensul dezvoltării unui fenomen, apoi pentru că formele de reprezentare sînt aici mult mai simple decît în repartițiile de frecvențe. De aceea, nu numai în lucrările de cercetare, ci și în popularizarea evoluției indicatorilor social-economici reprezentările grafice constituie o metodă curentă și generală de cunoaștere statistică. Formele caracteristice de reprezentare a seriilor cronologice sînt *cronograma* sau *historiograma* și *diagramele polare* și *în spirală*.

Cronograme sau historiograme se numesc construcțiile grafice tipice pentru seriile statistice cronologice; ele sînt analoage histogramelor utilizate, cum am văzut, la reprezentarea repartițiilor de frecvență și pentru construirea lor se folosesc coordonatele rectangulare, astfel cum se vede, de pildă, în figura 7.1. Aici apare evoluția a trei indicatori din tabelul 7.1, exprimați cu două feluri de unități de măsură, așa cum apar în anuarul statistic și, ca atare, au fost folosite două scări de gradație pe axa ordonatelor. Se poate folosi o singură scară, dar în acest caz și evoluția celui de-al treilea indicator trebuie să fie exprimat în cifre relative, cu baza în 1950. Fiind vorba de serii de intervale, punctele care marchează mărimea indicatorilor sînt plasate în mijlocul perioadelor. Acest procedeu se aplică și atunci cînd nivelurile sînt exprimate în mărimi medii.

Pentru a urmări comparativ evoluția a doi indicatori, de pildă a populației ocupate în agricultură în comparație cu creșterea populației ocupate în toate activitățile, se poate folosi modelul de reprezentare grafică din fig. 7.2. Fiind vorba de serii de momente, mărimile sînt plasate pe ordonatele ridicate pe axa abscisei cu scara timpului. Pentru a menține proporțiile desenului într-un spațiu economic și în limite estetice, gradația între 0 și 3000 a fost întreruptă.

Scanned with OKEN Scanner

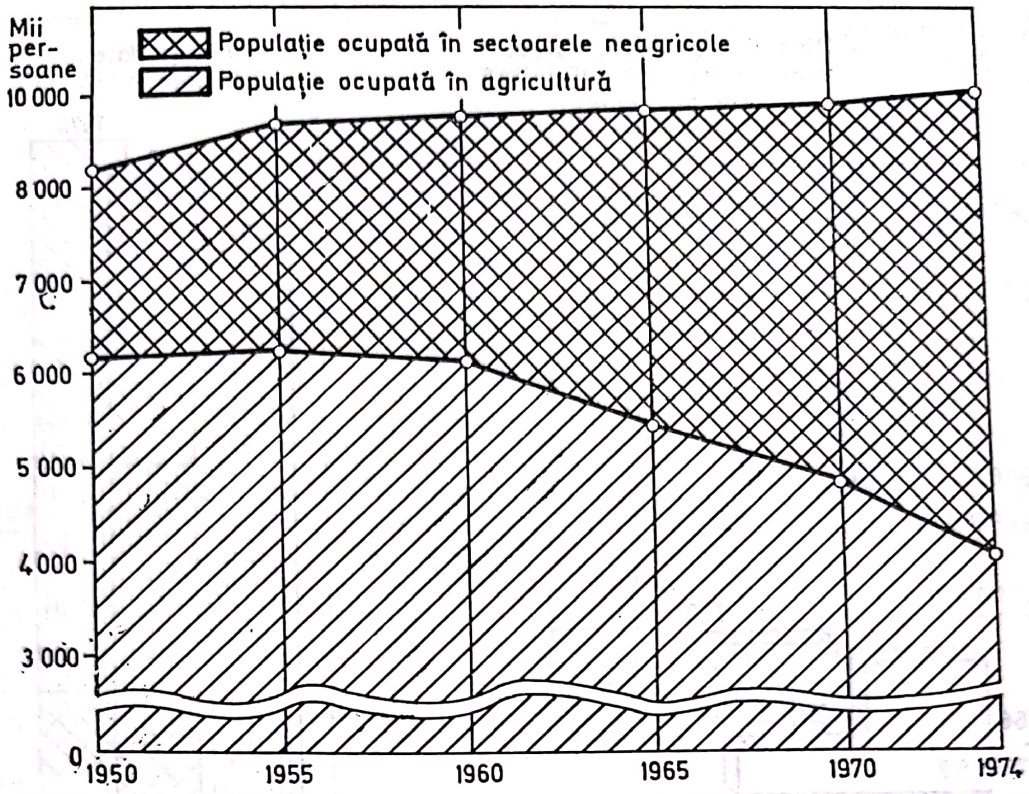


Fig. 7.2 Evoluția populației ocupate, în România, în anii 1950—1974 (mii persoane, la sfîrșitul anului). Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1975*.

Cînd se urmărește, comparativ, evoluția unor fenomene, dar între mărimile indicatorilor respectivi este o diferență foarte pronunțată, ceea ce îngreunează fixarea într-o reprezentare grafică, se recurge la scara semilogaritmică, precum se poate vedea în fig. 7.3.

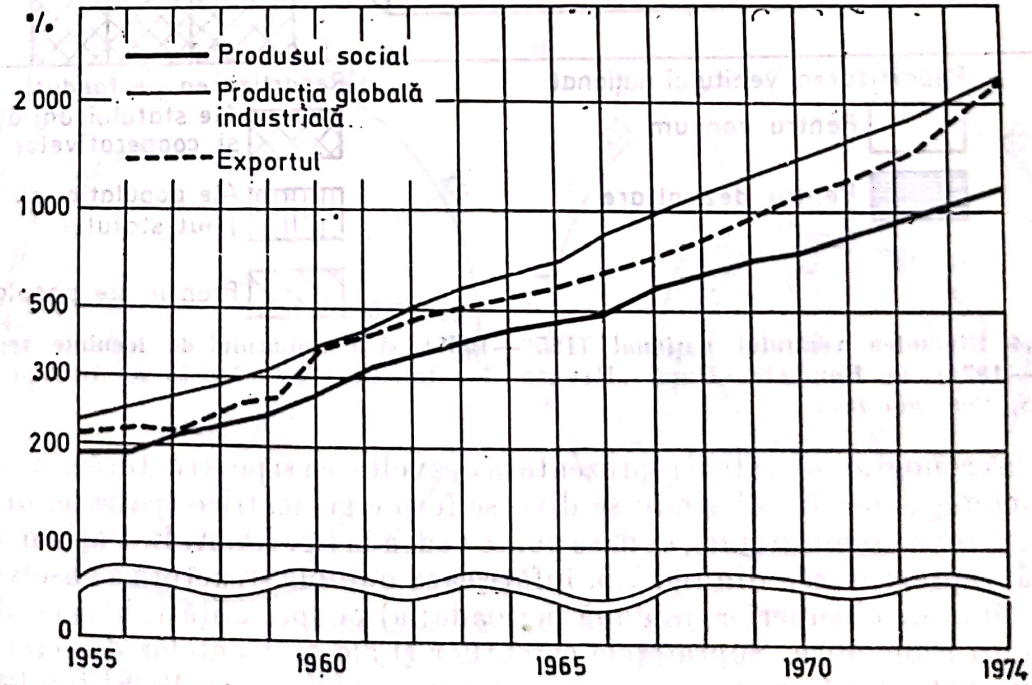


Fig. 7.3 Dinamica produsului social, a producției globale industriale și a exportului României în anii 1955—1974. (1950 = 100 scară logaritmică). Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1975*.

Scanned with OKEN Scanner

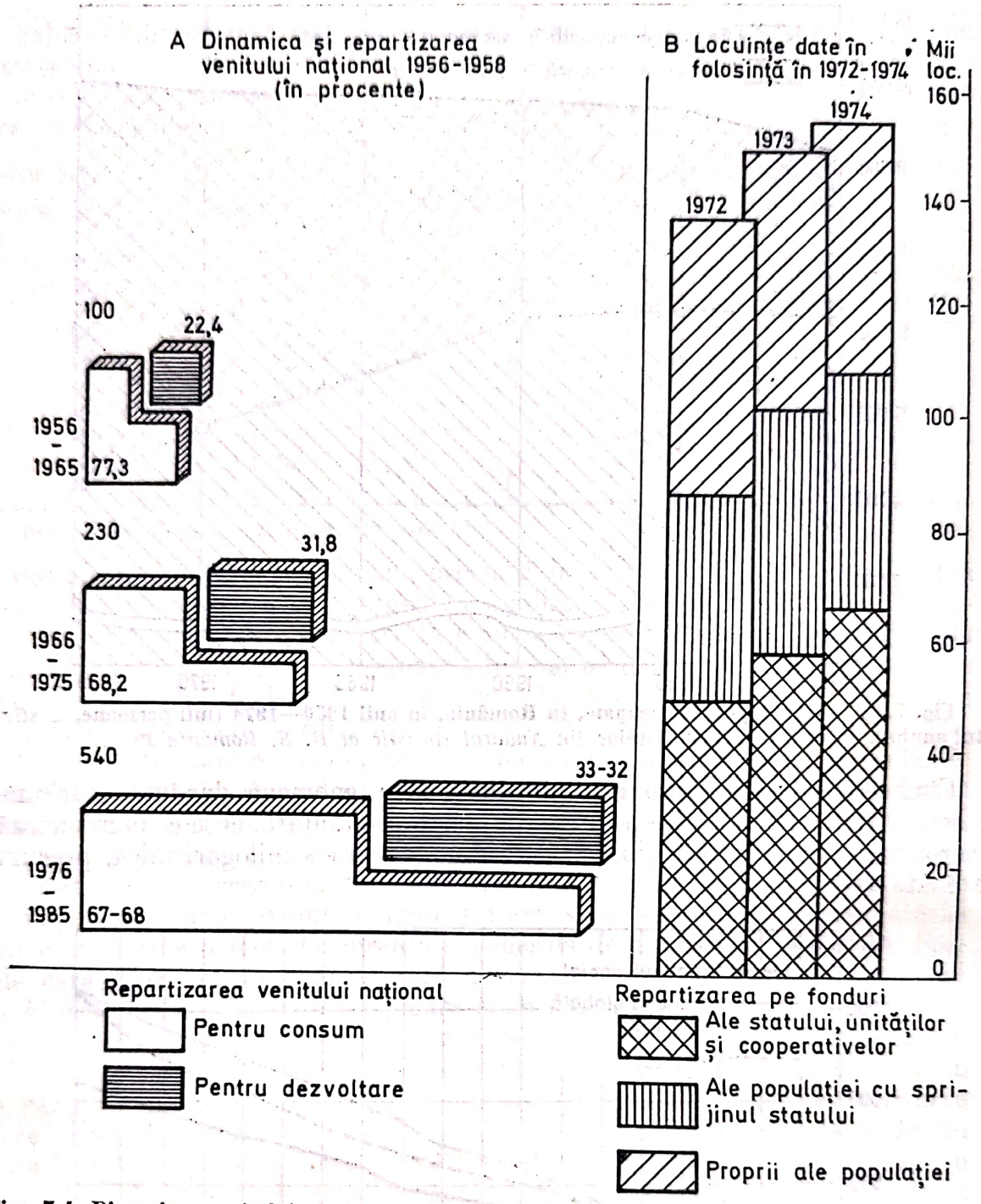


Fig. 7.4 Dinamica venitului național (1956—1985) și a numărului de locuințe repartizate (1972 — 1974) în România. După „Revista de statistică" nr. 7/1975 și *Anuarul statistic al R. S. România 1975.*

Prin cronograme poate fi reprezentată dezvoltarea și pe structură a indicatorilor economico-sociali, folosindu-se diverse forme geometrice (pătrate, dreptunghiuri, cercuri, semicercuri), cum se poate vedea în fig. 7.4. A, B. Puțin mai complicată, reprezentarea din fig. 7.5. înfățișează o dublă structură a absolvenților din învățământul superior, în două perioade: a) pe specialități; b) specialitățile pe ramuri economice. Suprafețele cercurilor și ale segmentelor de cercuri sînt proporționate după creșterea fenomenului între anii școlari 1938/39 și 1973/74.

În subcapitolele care urmează vor fi utilizate și alte forme de reprezentări grafice ale seriilor cronologice.

Scanned with OKEN Scanner

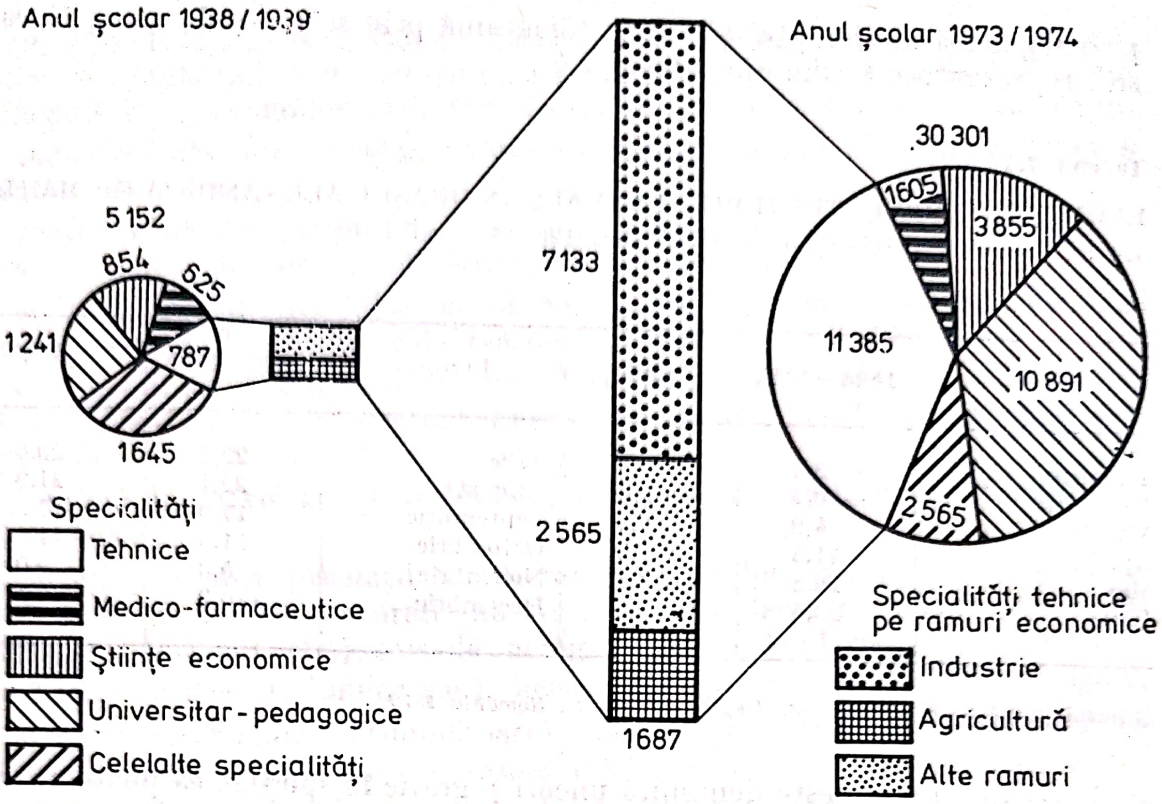


Fig. 7.5 Absolvenții din învățământul superior pe grupe de specialități, în România, în anii școlari 1938/39 și 1973/74. Sursa *Anuarul statistic al R. S. România 1975.*

Pentru urmărirea evoluției în timp se folosește, de obicei, diagrama polară sau radială, la care scara se construiește pe raza orizontală din dreapta (sau/și stînga, în sens invers), lungimea acesteia fiind egală cu media indicatorilor seriei reprezentate pe un ciclu, astfel cum se poate vedea în fig. 7.6 A (diagramă

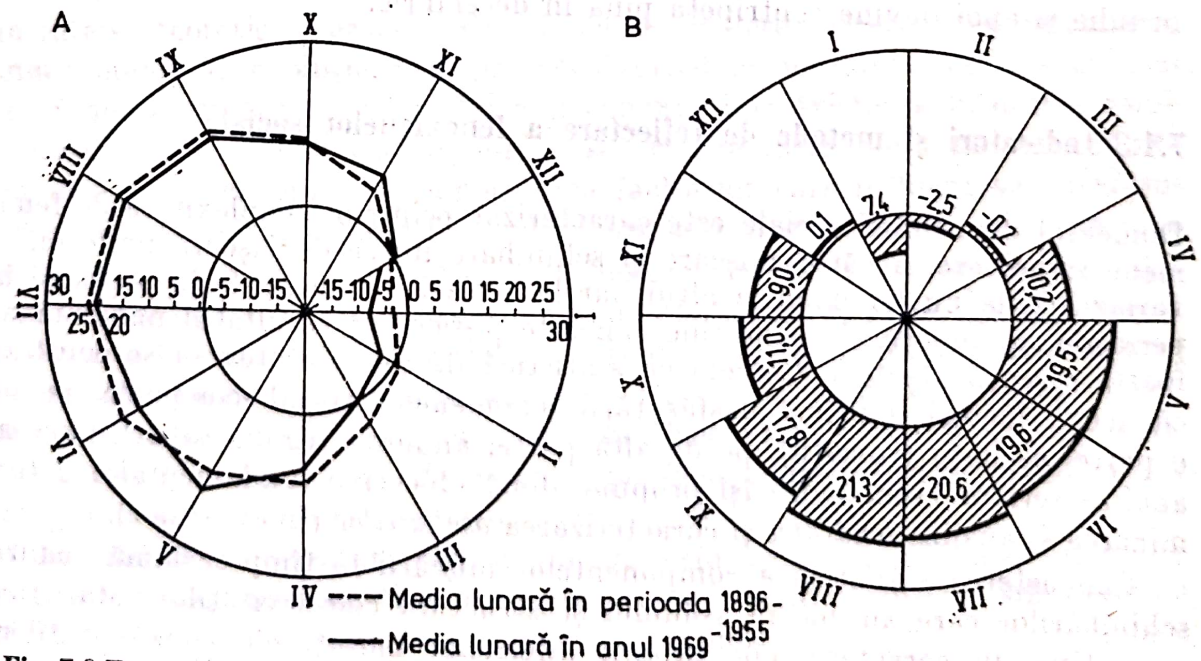


Fig. 7.6 Temperatura medie anuală o orașului Alexandria (România), în 1896—1955 (media anuală) și 1969.
Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1970.*

prin segmenţi de dreaptă) şi 7.6. B (diagramă prin sectoare de cerc), în care sînt reprezentate seriile din tabelul 7.4.

Tabelul 7.4

EVOLUŢIA TEMPERATURII MEDII ANUALE ÎN ORAŞUL ALEXANDRIA (ROMÂNIA), ÎN PERIOADA 1896—1955 ŞI ÎN ANUL 1969

(în grade Celsius)

Lunile	1896—1955	1969	Lunile	1896—1955	1969
Ianuarie	— 3,2	— 7,4	Iulie	22,7	20,6
Februarie	— 0,9	— 2,5	August	22,1	21,9
Martie	4,9	0,2	Septembrie	17,9	17,8
Aprilie	11,5	10,2	Octombrie	11,9	11,0
Mai	16,8	19,6	Noiembrie	5,3	9,0
Iunie	20,5	19,6	Decembrie	—0,2	0,1

Întocmit după datele din *Anuarul statistic al R. S. România 1970.*

Diagrama polară este denumită uneori şi grafic în spirală; ea poate apărea ca spirală centrifugă, dacă fenomenul este în creştere sau ca spirală centripetă, în cazul dinamicii regresive. De fapt, reprezentarea evoluţiei temperaturii în figura A ia forma unei spirale. Aceasta se observă dacă avem în vedere că, anticipînd datele din tabel cu luna decembrie, cînd media lunară pentru perioada 1896—1955 a fost de —0,2°, la fel cu cea din 1896—1956, deci punctele care marchează mărimea se acoperă, media lunară pe 1968 a marcat —1,4°, fiind deci mai coborîtă decît în 1969. Spirala este deci centrifugă pînă în iulie şi apoi devine centripetă pînă în decembrie.

7.1.3 Indicatori şi metode de reflectare a fenomenelor sociale

Domeniul dezvoltării sociale este caracterizat printr-o complexitate de fenomene şi procese, a căror mişcare şi schimbare nu se desfăşoară uniform, ci variază de la un fenomen la altul, iar la acelaşi fenomen, de la perioadă la perioadă, în funcţie de condiţiile concrete, reflectînd rezultatul unui număr foarte mare de factori. Cercetarea statistică dă posibilitatea să se constate că aceşti factori imprimă desfăşurării fenomenelor social-economice, pe de o parte, oscilaţii (variaţii), pe de altă parte, anumite regularităţi. De aceea, analiza seriilor cronologice îşi propune două obiective fundamentale: determinarea şi analiza *tendinţei* şi caracterizarea *oscilaţiilor* (fluctuaţiilor).

Cunoaşterea statistică a componentelor mişcării în timp reclamă analiza schimbărilor care au loc în volumul şi structura colectivităţilor statistice respective, în caracteristicile proprii diferitelor unităţi ale colectivităţilor, în forma şi caracterul legăturii dintre aceste caracteristici, în toate schimbările manifestîndu-se relaţii de interdependenţă. [85].

Scanned with OKEN Scanner

Plecînd de la expresiile valoare absolută, valoare medie şi valoare relativă, indicatorii statistici prin care sînt caracterizate mişcările şi schimbările din domeniul social-economic reflectate în serii cronologice sînt şi ei clasificaţi în:

indicatori absoluţi: nivelul, volumul şi sporul absolut al seriei, precum şi volumul sporului;

indicatori relativi: ritmul de creştere şi ritmul sporului; de fapt sînt indicatorii cunoscuţi, de obicei, sub denumirea de indici statistici;

indicatori medii: nivelul şi sporul mediu al seriei, ritmul mediu de creştere şi ritmul mediu al sporului, valoarea medie a unui procent al sporului.

Ca efect al acţiunii factorilor care determină modul de dezvoltare (evoluţie) a fenomenelor oglindite în seriile dinamice, mişcările şi schimbările apar sub formă de:

tendinţă de dezvoltare pe o durată mai lungă, marcînd direcţia fundamentală a mişcării;

variaţii periodice (oscilaţii sezoniere), care au loc în anumite condiţii şi sînt create de factori naturali (anotimp, sezon), factori cu caracter social (concedii, sărbători etc.) sau de neregularităţi calendaristice (inegalitatea numărului de zile ale lunilor ş.a.); aceste variaţii au o periodicitate scurtă;

oscilaţii incidentale (întîmplătoare), numite şi neregulate, care apar ca rezultat al acţiunii unor factori întîmplători, precum sînt calamităţile naturale, războaiele, măsurile excepţionale cu caracter politic, economic, administrativ etc.; aceste variaţii sînt numite uneori *reziduale.*

La cercetarea statistică a fenomenelor din ţările capitaliste este necesar să avem în vedere şi *variaţiile ciclice*, determinate de periodicitatea succesiunii fazelor de ciclu în procesul reproducţiei din economia capitalistă.

Analiza statistică a seriilor cronologice impune disocierea componentelor seriei, izolarea diverselor moduri de manifestare a mişcării, aceasta prezentînd un interes teoretic şi practic. Din punct de vedere teoretic, descompunerea unui fenomen în evoluţie, prin procedee statistice, permite o cunoaştere mai aprofundată, stabilirea influenţelor şi conexiunilor existente în desfăşurarea fenomenului, a interdependenţelor ce se creează între două sau mai multe fenomene. Pentru practică, cunoaşterea factorilor care influenţează desfăşurarea fenomenelor şi proceselor, a legăturilor dintre ei, a abaterilor de la o desfăşurare normală (planificată) este desigur de cea mai mare importanţă. Cunoaşterea oscilaţiilor (perturbaţiilor) care afectează, de pildă, aprovizionarea planificată cu materii şi materiale, efectuarea transporturilor necesare, productivitatea muncii, mai ales în condiţiile activităţii complexe a întreprinderilor moderne, dă posibilitatea conducerii, la diverse niveluri ca, pe calea conexiunii inverse, să ia măsuri de aducere pe făgaşul desfăşurării planificate a procesului afectat de perturbaţii.

În lumina celor arătate pînă acum, în desfăşurarea fenomenelor oglindite în seriile cronologice pot fi identificate anumite laturi (aspecte) sub care ele pot fi privite şi, ca atare, cercetarea presupune:

— studierea seriei cu ajutorul indicatorilor absoluţi şi relativi, precum şi al nivelului mediu;

Scanned with OKEN Scanner

— utilizarea teoriei indicilor statistici, pentru determinarea dezvoltării de la o perioadă la alta și a influenței factorilor;

— cunoașterea dezvoltării generale prin determinarea liniei de tendință (trendul);

— măsurarea oscilațiilor sezoniere.

Corespunzător acestor aspecte se aplică indicatorii aferenți și procedee anumite de prelucrare tehnică.

7.2. INDICATORI ABSOLUȚI ȘI RELATIVI. CALCULUL RITMULUI MEDIU DE DEZVOLTARE

7.2.1 Indicatori absoluți și relativi

Nivelul absolut al seriei, fie că este vorba de o serie de momente sau de una de intervale, este reprezentat prin valoarea absolută a fiecărui termen din serie, notat în calculele statistice, de pildă, cu $y_1, y_2, \ldots, y_n$. Însumarea seriei de intervale este concretizată în indicatorul *volumului absolut* al acesteia, adică:

$$y_1 + y_2 + \ldots + y_n = \Sigma y.$$

Diferența dintre doi termeni ai seriei, care apar ca mărimi absolute, se exprimă prin indicatorul *sporul absolut* (creșterea absolută) și el ia *forma de bază*, cînd diferența se calculează între unul din termeni și termenul inițial, adică

$$\Delta_{n/_0} = y_n - y_0$$

sau *în lanț*, în care caz diferența se determină între doi termeni ce se succed, conform formulei:

$$\Delta_{n/_{n-1}} = y_n - y_{n-1}.$$

Pentru seriile de intervale se calculează și *volumul sporului*, prin însumarea sporurilor în lanț sau reprezentînd sporul în formă de bază a ultimei perioade, adică:

$$\Delta_{1/_0} + \Delta_{2/_1} + \ldots + \Delta_{n/_{n-1}} = \Sigma\Delta = \Delta_{n/_0}.$$

Aplicarea mărimilor relative în cercetarea statistică a seriilor dinamice se caracterizează mai întîi în indicatorul *ritmului de creștere* [1], care se determină, în formă de bază:

$$I_{n/_0} = \frac{y_n}{y_0}, \quad \text{respectiv} \quad I_{n/_0} = \frac{y_n}{y_0} \cdot 100,$$

prin raportarea fiecărui termen al seriei la unul de bază, de obicei termenul inițial.

[1] Mai corect ar fi, al *ratei de creștere*.

Scanned with OKEN Scanner

$I_{n/_0}$ este denumit *indice* sau ritm (rată) de creştere şi se exprimă, de obicei, în procente. El poate fi determinat şi în lanţ, prin raportarea fiecărui termen al seriei la cel precedent, adică

$$I_{n/n-1} = \frac{y_n}{y_{n-1}}, \quad \text{respectiv} \quad I_{n/n-1} = \frac{y_n}{y_{n-1}} \cdot 100.$$

Ritmul sporului, ca mărime relativă de dinamică, se calculează, de asemenea, în formă de bază şi în lanţ, cu ajutorul formulelor;

$$R_b = \frac{\Delta_{n/_0}}{y_0} \quad \text{sau} \quad R = \frac{y_n - y_0}{y_0} = \frac{y_n}{y_0} - 1, \quad \text{respectiv} \quad \frac{y_n}{y_0} \cdot 100 - 100$$

$$R_l = \frac{y_{n/n-1}}{y_{n-1}} \quad \text{sau} \quad \frac{y_n - y_{n-1}}{y_{n-1}} = \frac{y_n}{y_{n-1}} - 1, \quad \text{respectiv} \quad \frac{y_n}{y_{n-1}} \cdot 100 - 100.$$

Indicatorii medii. Pentru calcularea indicatorilor medii ai seriilor cronologice se folosesc, de obicei, următoarele formule:

La determinarea *nivelului mediu*, se aplică formula mediei aritmetice $\bar{y} = \frac{\Sigma y}{n}$, pentru seriile cronologice de intervale şi formula mediei cronologice, pentru seriile cronologice de momente, adică:

$$\bar{y} = \frac{^1/_2\, y_1 + y_2 + \ldots + {}^1/_2\, y_n}{n-1}, \quad \text{formulă care se mai poate scrie:}$$

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_2}{2} + \ldots + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{^1/_2\, y_1 + y_2 + \ldots + y_{n-1} + {}^1/_2\, y_n}{n-1}.$$

Sporul absolut mediu se calculează pe baza sporurilor absolute în formă de bază şi în lanţ, pentru seriile de momente şi de intervale, cu ajutorul formulelor

$$\bar{\Delta}_b = \frac{\Sigma\Delta}{n-1}; \quad \bar{\Delta}_l = \frac{\Delta_{n/_0}}{n-1}.$$

Ritmul mediu de creştere, care reprezintă media ritmurilor de creştere în lanţ ale seriei, se determină cu ajutorul unor metode şi procedee diferenţiate în funcţie de particularităţile fenomenelor cercetate şi de obiectivele urmărite de cercetarea statistică. Aceste metode vor fi dezvoltate în paragraful următor.

Ritmul mediu al sporului se calculează pe baza indicatorului precedent, cu ajutorul formulei:

$$\bar{r} = I_g - 1 \quad \text{sau} \quad \bar{r} = I_g\, 100 - 100.$$

Cu ajutorul unui tabel de calcul putem prezenta, în mod concentrat şi sinoptic, indicatorii absoluţi, relativi şi medii ai seriilor cronologice, folosiţi

Scanned with OKEN Scanner

pentru oglindirea dezvoltării exportului țării noastre, pe o perioadă de 11 ani (tabelul 7.5).

Tabelul 7.5

DINAMICA EXPORTULUI ROMÂNIEI ÎN PERIOADA 1959—1969

Ani	Mld lei valută y	Sporul absolut (mld lei val.)		Ritmul de creștere (%)		Ritmul sporului (%)		Nivelul mediu (mld. lei val.)	Sporul mediu anual (mld. lei val.)	Ritmul mediu de creștere (%) (1959 — 1969)
		bază	lanț	bază	lanț	bază	lanț			
		$y_n - y_0$	$y_n - y_{n-1}$	$\frac{y_0}{y_n} \cdot 100$	$\frac{y_n}{y_{n-1}} \cdot 100$	$\frac{y_n}{y_0} 100 - 100$	$\frac{y_n}{y_{n-1}} 100 - 100$	$\frac{\Sigma y}{n}$	$\frac{\Sigma \Delta}{n-1}$	$r = \sqrt[10]{I_g} - 1$
1959	3,1	—	—	100,0	—	—	—			
1960	4,3	1,2	1,2	138,7	138,7	38,7	38,7			
1961	4,8	1,7	0,5	154,8	116,6	54,8	11,6			
1962	4,9	1,8	0,1	158,1	102,1	58,1	2,1			
1963	5,5	2,4	0,6	177,4	112,2	77,4	12,2			
1964	6,0	2,9	0,5	193,5	109,1	93,5	9,1	6,3	0,66	12,07
1965	6,6	3,5	0,6	213,0	110,0	113,0	10,0			
1966	7,1	4,0	0,4	229,0	107,6	129,0	7,6			
1967	8,4	5,3	1,3	271,0	118,3	171,0	18,3			
1968	8,8	5,7	0,4	284,0	104,8	184,0	4,8			
1969	9,8	6,7	1,0	316,1	111,4	216,1	11,4			

Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1960* și *1975*.

Pentru acest exemplu, ritmul mediu de creștere a fost determinat pe baza mediei geometrice, urmărindu-se simplificarea calculului și considerîndu-se că linia de dezvoltare a fenomenului permite aplicarea acestei metode. Întrucît realitatea oferă și alte caracteristici, ritmul mediu de dezvoltare poate fi calculat și cu ajutorul altor metode, pe care le vom prezenta în cele ce urmează.

7.2.2 Metode de determinare a ritmului mediu

Aplicarea mediei geometrice și a mediei parabolice. Utilizată în mod curent în foarte diverse scopuri practice — ca element de calcul pentru cunoașterea desfășurării unui fenomen, a unei activități, a abaterilor de la o evoluție unidirecțională, pentru comparații între diferite perioade ale dezvoltării social-economice, sub aspectul indicatorilor pertinenți, pentru preliminarii și calcule de perspectivă — metodologia de determinare a ritmului mediu de creștere s-a găsit în fața unor probleme dificile.

În general, se consideră că un fenomen se dezvoltă ritmic atunci cînd mărimile relative în lanț sau sporurile relative ale valorilor succesive ale seriei sînt aproximativ egale, ceea ce justifică stabilirea ritmului mediu numai

Scanned with OKEN Scanner

pentru seriile cronologice cu creşteri sau descreşteri, care prezintă o anumită constanţă. Metodologia de calcul s-a diferenţiat oarecum într-un cadru destul de îngust, urmărindu-se, în principal, construirea unui indicator, care să fie sensibil la toate oscilaţiile seriilor cu tendinţă evidentă de creştere sau descreştere. Pentru celelalte serii, diverşi autori au considerat că orice soluţie este formală şi lipsită de sens. Prin însăşi definiţia sa, ritmul presupune o dezvoltare treptată a fenomenului, în ritm uniform, fiind deci caracteristic seriilor ce pot fi aproximate prin funcţii exponenţiale. Pentru seriile care, din punct de vedere matematic, corespund unor funcţii parabolice de diferite ordine sau hiperbolice, nu este indicat să se calculeze ritmul mediu. În orice caz aplicarea metodologiei de calcul al ritmului mediu trebuie să se facă cu mult discernămînt, evitîndu-se criteriile formale şi studiindu-se manifestarea concretă a tendinţei de dezvoltare a fenomenului [73].

Dacă urmărim evoluţia seriei cronologice din tabelul 7.5., aşa cum ne arată cifrele sporului şi ritmului de creştere (în formă de bază), constatăm că direcţia evoluţiei păstrează o anumită constanţă de creştere. Aceasta se observă mai bine într-o reprezentare grafică a datelor empirice (figura 7.3). Realitatea construcţiei socialiste de la noi şi din alte ţări socialiste demonstrează că şi alţi indicatori sintetici social-economici urmează aceeaşi linie de mişcare.

Într-o asemenea aplicaţie se justifică utilizarea *mediei geometrice*, care se bazează pe relaţia dintre termenul final al seriei şi produsul termenilor unui şir de valori care reprezintă coeficienţi de creştere (k), iar relaţia se scrie:

$$y_n = y_0 \cdot k_1 \cdot k_2 \dots k_n,$$

ea fiind satisfăcută atunci cînd

$$y_0 \cdot k_1 \cdot k_2 \dots k_n = y_0 \bar{k}^n.$$

Simplificînd ecuaţia cu y_0 şi extrăgînd rădăcina de ordinul n, obţinem valoarea medie a coeficienţilor de creştere, adică ritmul mediu al creşterii

$$\bar{k} = \sqrt[n]{k_1 \cdot k_2 \dots k_n}.$$

Produsul coeficienţilor în lanţ se reduce la coeficientul în formă de bază al ultimei perioade, care conform notaţiilor anterioare este $I_{n/0}$, adică:

$$\bar{I}_g = \sqrt[n-1]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \dots \frac{y_n}{y_{n-1}}} = \frac{y_n}{y_0} \quad \text{sau} \quad \sqrt[n-1]{I_{n/0}}.$$

Folosind logaritmii, relaţia devine

$$\log \bar{I}_g = \frac{(\log y_1 - \log y_0) + (\log y_2 - \log y_0) + (\log y_n - \log y_{n-1})}{n-1} =$$

$$= \frac{\log y_n - \log y_0}{n-1} = \frac{\log J_{n/0}}{n-1}$$

şi apoi prin antilogaritmare obţinem tocmai valoarea indicatorului ritmului mediu de creştere.

Plecînd de la ipoteza că o dezvoltare continuu crescîndă (coborîtoare) nu exclude şi unele abateri [85], cercetările aplicative se prevalează, obişnuit la determinarea ritmului mediu, de metoda mediei geometrice. S-a

Scanned with OKEN Scanner

obiectat însă că această medie este sensibilă numai la primul şi ultimul termen al seriei şi ignorează termenii intermediari, astfel că ea nu poate fi aplicată mai ales la evoluţii oscilatorii, cum ar fi de pildă seriile din reprezentarea grafică.

În consecinţă, în anul 1949, matematicianul român Al. Froda [2] a recomandat aplicarea metodei *mediei parabolice*, care ţine seama de suma valorilor termenilor seriei şi de valoarea termenului iniţial, pornind de la relaţia:

$$y_0 + y_0\theta + y_0\theta^2 + y_0\theta^3 + \ldots + y_0\theta = \Sigma y_i,$$

în care y_0 reprezintă termenul iniţial al seriei şi n — numărul termenilor.

Scoţînd pe y_0 factor comun şi dezvoltînd calculul, obţinem ecuaţia:

$$1 + \theta + \theta^2 + \ldots + \theta^n = \frac{\Sigma y_i}{y_0},$$

aici prima parte reprezentînd suma termenilor unei progresii geometrice, astfel că putem scrie şi sub forma

$$\frac{\theta^{n+1} - 1}{\theta - 1} = \frac{\Sigma y_i}{y_0}.$$

Obiecţiile aduse metodei mediei parabolice se referă, în primul rînd, la faptul că, întrucît calculul este orientat după suma nivelurilor seriei, ca şi după nivelul primului termen, preferinţa acordată acestuia nu este justificată. Dacă între primul termen şi următorii sînt diferenţe mari, ritmul mediu este influenţat în sus sau în jos. O obiecţie, din punct de vedere teoretic, s-a făcut în sensul că dacă modelul matematic reprezintă o funcţie pătratică şi curba este o parabolă de gradul doi, prima derivată şi raportul acesteia faţă de funcţia respectivă sînt mărimi variabile. În consecinţă, atunci cînd schimbările au ca model o parabolă de gradul doi, calculul ritmului nu are sens, deoarece lipseşte ritmul constant apropiat de ritmul mediu. O deficienţă vizibilă a acestei metode, proprie dealtfel şi metodei geometrice, constă în faptul că ea desprinde doar cîteva faze din evoluţia fenomenului oglindit în seria cronologică. Ca atare, nici una din aceste medii nu este utilizabilă, dacă seria nu reprezintă o dezvoltare uniformă, în sens absolut sau relativ. Ambele însă pot fi aplicate la serii caracteristice printr-o relativă constanţă a dezvoltării, fiind în linii mari echivalente [85].

A treia metodă de determinare a ritmului mediu de creştere se bazează pe utilizarea *funcţiei exponenţiale* şi a *metodei celor mai mici pătrate* şi a fost recomandată de acad. Gh. Mihoc şi de V. Urseanu [85], precum şi de I. Marinescu [73].

Aplicarea funcţiei exponenţiale se justifică în cazul existenţei unei creşteri relativ uniforme a valorilor unei serii, care poate fi exprimată printr-o funcţie exponenţială de forma

$$y = ab^x,$$

în care

y reprezintă valorile termenilor seriei;
x — timpul;
a şi b — valori constante.

[2] *Al. Froda, Numere-indice cumulatoare*, „Analele Academiei R.P.R.", III, 1949.

Scanned with OKEN Scanner

Aici a indică punctul de intersecție a curbei cu axa ordonatelor, iar b gradul de înclinație a curbei — panta sau viteza ei —, adică tocmai ritmul mediu de creștere. Prin logaritmarea funcției exponențiale de mai sus se obține

$$\log y = \log a + x \log b$$

sau

$$y' = a' + xb',$$

care este o funcție liniară a logaritmilor.

În general, creșterea poate fi descrisă cu ajutorul progresiei aritmetice și al progresiei geometrice.

Progresia aritmetică este reprezentată de o dreaptă:

$$y = a + bx,$$

formulă cu care ne vom mai întîlni la analiza tendinței și la ajustări și neteziri.

Progresia geometrică este de fapt o curbă exponențială dată de ecuația:

$$y = ab^x,$$

prin logaritmare: $\log y = \log a + x \log b$.

Cea mai familiară expresie a acestei ecuații este

$$P_n = P_0 (1 + r)^n,$$

care este ecuația dobînzii compuse.

Prin logaritmare, ea devine:

$$\log P_n = \log P_0 + n \log (1 + r).$$

Forma cea mai generală a funcției exponențiale este:

$$P_n = P_0 \cdot e^{rn},$$

în care:

e reprezintă baza logaritmilor naturali (2, 71828 ...)

$\log e = M = 0{,}43429 \ldots$

De aici:

$$e^{rn} = \frac{P_n}{P_0}.$$

Prin logaritmare:

$$r = \frac{\log \left(\frac{P_n}{P_0} \right)}{n \log e}.$$

Exemplu: Venitul național pe un locuitor în România a fost de 7670 lei în 1965 și 17 000 în 1975.

Scanned with OKEN Scanner

Calculul ritmului (ratei) mediu de creştere în perioada 1965—1975:

a) cu ajutorul ecuaţiei dobînzii compuse (mediei geometrice):

$$(1 + r)^{10} = \frac{V\,1975}{V\,1965}; \quad (1 + r) = \sqrt[10]{\frac{V\,1975}{V\,1965}};$$

$$r = \sqrt[10]{\frac{V\,1975}{V\,1965}} - 1 \cdot (1 + r) = \sqrt[10]{\frac{17\,000}{7\,670}} = (1 + r) = \sqrt[10]{2{,}2164}$$

$$\log (1 + r) = \frac{\log 2{,}2164}{10} = \frac{0{,}34565}{10} = 0{,}034565$$

antilogaritm $0{,}034565 = 1{,}083$ sau $108{,}3\%$

$r = 1{,}083 - 1 = 0{,}083$ sau $8{,}3\%$.

b) cu ajutorul funcţiei exponenţiale:

$$e^{rn} = \frac{17\,000}{7\,670}$$

$$r = \frac{\log\left(\frac{17000}{7\,670}\right)}{n \log e} = \frac{0{,}3456}{10 \cdot 0{,}43429} = 0{,}0796$$

sau $7{,}96\% \approx 8\%$.

În general, funcţia exponenţială dă valori mai mici decît media geometrică.

În mod obişnuit se aplică media geometrică. În anuarele O.N.U., F.A.O, B.I.T., ca şi în anuarele statistice naţionale, calculul ratei medii de evoluţie a fenomenelor (populaţia, producţia etc.) se face cu ajutorul mediei geometrice.

Metoda celor mai mici pătrate. Pentru calcularea parametrilor a şi b, pe baza cunoaşterii tipului funcţiei şi a valorii termenilor seriei, se foloseşte metoda celor mai mici pătrate; întrucît aceasta va fi utilizată şi la determinarea liniei de tendinţă, o prezentăm aici mai întîi în elementele şi caracteristicile sale fundamentale, urmînd a o aplica la calcularea ritmului mediu de creştere, ca metodă a trendului exponenţial.

Metoda celor mai mici pătrate a fost formulată de K.Fr. Gauss [3] şi utilizată în calculul compensării abaterilor pe baza principiului care, în analiza seriilor cronologice, s-ar putea defini astfel: suma pătratelor diferenţei $(y - Y)^2$ dintre y — mărimile reale ale termenilor seriei — şi Y — mărimile obţinute prin calcul (ordonatele corespunzătoare punctelor situate pe linia de tendinţă), respectiv $\sum (y - Y)^2$ — trebuie să fie minimă. Concret, prin această metodă se urmăreşte găsirea ecuaţiei liniei drepte sau curbe, care să înlocuiască datele reale ale seriei dinamice, satisfăcînd principiul enunţat.

[3] KARL FRIEDRICH GAUSS (1777—1855), astronom, matematician şi fizician german, autorul unor lucrări de mecanică astrală, magnetism şi optică, a adus contribuţii remarcabile la dezvoltarea calculului probabilităţilor, mai ales prin formularea legii distribuţiei erorilor (abaterilor) ca o curbă în formă de clopot („clopotul lui Gauss").

Scanned with OKEN Scanner

Considerînd că ecuaţia unei curbe este dată de $y = f(x)$, în care mărimile y sînt funcţii de x, vom utiliza, după caz, ecuaţia dreptei, a parabolei, a hiperbolei, a funcţiei exponenţiale sau a unui polinom de un grad oarecare.

Tipul ecuaţiei (funcţiei) urmează a fi ales de statistician, care are a decide, ţinînd seama de natura fenomenului cercetat, de datele reale ale seriei cronologice şi de alţi factori.

În expresia $\sum_{i=1}^{n} (y_i - Y_i)^2 = \text{minim}$, $Y = \log y_i$, adică logaritmii valorilor reale ale seriei, ea poate fi privită ca o funcţie a parametrilor a şi b, adică $f(a, b)$ şi este minimă cînd derivatele parţiale ale funcţiei în raport cu a şi b sînt egale cu zero, adică

$$\frac{\delta\varphi}{\delta a} = 2\sum_{i=1}^{n}(a' + b'\,x_i - Y_i) = 0$$

$$\frac{\delta\varphi}{\delta b} = 2\sum_{i=1}^{n}(a' + b'\,x_i - Y)\,x_i = 0.$$

Să presupunem că a fost aleasă funcţia dreptei (liniară) de gradul I, cu formula:

$$y = a + bx,$$

în care:

x sînt mărimile date timpului (anilor) din seria dinamică;
y — mărimile reale ale termenilor seriei;
a şi b — coeficienţi (parametri), pe care-i calculăm prin metoda celor mai mici pătrate şi care indică direcţia şi schimbarea tendinţei.

Astfel, obţinem relaţiile:

$$y_1 = a + bx_1$$
$$y_2 = a + bx_2$$
$$\cdots\cdots\cdots\cdots$$
$$y_n = a + bx_n$$
$$\overline{\Sigma y = na + b\Sigma x.}$$

A doua ecuaţie o obţinem înmulţind-o pe prima cu x, adică

$$xy = ax + bx^2,$$

în care, dînd lui x şi y mărimile corespunzătoare, vom avea:

$$x_1y_1 = ax_1 + bx_1^2$$
$$x_2y_2 = ax_2 + bx_2^2$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_ny_n = ax_n + bx_n^2$$
$$\overline{\Sigma xy = a\,\Sigma x + b\Sigma x^2.}$$

Scanned with OKEN Scanner

Astfel, am obținut sistemul

$$\Sigma y = na + b\Sigma x$$
$$\Sigma xy = n\Sigma x + b\Sigma x^2,$$

care, rezolvat în raport cu a și b, și cu ajutorul determinanților, ne dă:

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma x \\ \Sigma xy & \Sigma x^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{\Sigma x^2 \cdot \Sigma y - \Sigma x \cdot \Sigma xy}{n \cdot \Sigma x^2 - (\Sigma x)^2}$$

$$b = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma x & \Sigma xy \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{n \cdot \Sigma xy - \Sigma x \cdot \Sigma y}{n \cdot \Sigma x^2 - (\Sigma x)^2}.$$

Calculul de mai sus poate fi simplificat. Axa Y, care reprezintă ordonata tendinței și este determinată de mărimile calculate, este aleasă astfel încît originea să cadă în mijlocul seriei, ceea ce face ca mărimile ei să se distribuie în sus și în jos, unele cu semnul minus (—), celelalte cu plus (+), iar la însumare ele să se anuleze reciproc, adică

$$\Sigma x = 0.$$

În această condiție, formulele pentru a și b se simplifică, în sensul că

$$a = \frac{\Sigma y}{n}, \quad \text{iar} \quad b = \frac{\Sigma xy}{\Sigma x^2}.$$

În ipoteza că funcția aleasă este parabola, cu ecuația

$$y = a + bx + cx^2,$$

trebuie să obținem un sistem de trei ecuații, din care să se calculeze parametrii a, b, c și care, după procedeul arătat, va fi

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4.$$

Alegînd, pentru simplificarea calculelor, axa y astfel încît originea să cadă în mijlocul seriei, deci

$$\Sigma x = 0; \quad \Sigma x^3 = 0,$$

ecuațiile de mai sus se simplifică și se obține sistemul:

$$\Sigma y = n \cdot a + c\Sigma x^2$$
$$\Sigma xy = b\Sigma x^2$$
$$\Sigma x^2 y = a\Sigma x^2 + c\Sigma x^4.$$

Din ecuația a doua obținem

$$b = \frac{\Sigma xy}{\Sigma x^2},$$

Scanned with OKEN Scanner

iar sistemul format din ecuaţia 1 şi a 3-a ne conduce la

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma x^2 \\ \Sigma x^2 y & \Sigma x^4 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x^2 \\ \Sigma x^2 & \Sigma x^4 \end{vmatrix}} = \frac{\Sigma x^4 \cdot \Sigma y - \Sigma x^2 \cdot \Sigma x^2 y}{n \cdot \Sigma x^4 - (\Sigma x^2)^2}$$

$$c = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma x^2 & \Sigma x^2 y \end{vmatrix}}{\begin{vmatrix} n & \Sigma x^2 \\ \Sigma x^2 & \Sigma x^4 \end{vmatrix}} = \frac{n \cdot \Sigma x^2 y - \Sigma x^2 \cdot \Sigma y}{n \cdot \Sigma x^4 - (\Sigma x^2)^2}.$$

Practic s-a dovedit că în cele mai multe cazuri este suficientă funcţia liniară

$$y = a + bx.$$

Dacă o serie dinamică va arăta o constanţă a creşterii (descreşterii) absolute de la un termen la altul, statisticianul va alege funcţia liniară, cu ecuaţia de gradul I arătată mai sus. Dacă însă dinamica seriei va arăta o constanţă a creşterii (descreşterii) relative, se va alege funcţia exponenţială, cu ecuaţia

$$y = a\, b^x,$$

care, prin logaritmare, ne dă o ecuaţie liniară:

$$\log y = \log a + x \log b.$$

La aplicarea ecuaţiei exponenţiale, pe baza metodei celor mai mici pătrate, în calculul ritmului mediu, plecăm de la ultimele formule, în care:

$$\log a = \frac{\Sigma \log y}{n} \quad \text{şi} \quad \log b = \frac{\Sigma x \log y}{\Sigma x^2}$$

şi folosim aici exemplificarea dată de I. Marinescu în studiul citat privind operaţiile de calcul, astfel cum se prezintă în tabelul 7.6.

Tabelul 7.6

SUPRAFAŢA ARABILĂ ÎN SECTORUL COOPERATIST DIN ROMÂNIA, ÎN PERIOADA 1950–1959
(la începutul anului)

Ani	Suprafaţa arabilă (mii ha)	Coeficientul ritmului de creştere	Valori teoretice ale seriei calculate prin		
			media geometrică	media parabolică	metoda exponenţială
1950	13,1	–	13,1	13,1	81,3
1951	262,5	20,030	24,8	26,2	130,2
1952	274,4	1,045	46,9	52,8	209,2
1953	851,3	3,102	88,7	106,1	336,2
1954	961,2	1,129	167,8	213,2	540,3
1955	1075,2	1,118	317,4	428,3	868,2
1956	1179,8	1,097	600,6	860,4	1395,3
1957	1691,0	1,433	1136,4	1728,5	2242,1
1958	3365,3	1,990	2150,2	3472,6	3603,1
1959	4071,9	1,209	4068,4	6976,5	5790,4

Scanned with OKEN Scanner

Aplicarea procedeelor de calcul pe baza celor trei metode ne conduce la rezultate sensibil diferite în ce privește valoarea ritmului mediu de creștere. Pentru calcularea trendului exponențial ne servim de următorul tabel ajutător:

Ani	Suprafața arabilă (mii ha) y	x	x^2	$\log y$	$x \log y$
1950	13,1	−9	81	1,1172713	−10,0555517
1951	262,3	−7	49	2,4191293	−16,9339051
1952	274,4	−5	25	2,4383841	−12,1919205
1953	851,3	−3	9	2,9300826	− 8,7902478
1954	961,2	−1	4	2,9828138	− 2,9828138
1955	1 075,2	1	4	3,0314893	3,0314893
1956	1 179,8	3	9	3,0718084	9,2154252
1957	1 691,0	5	25	3 2281436	16,1407180
1958	3 365,3	7	49	3,5271141	24,6897987
1959	4 071,9	9	81	3,6097971	32,4881739
Total	13 745,7	—	336	28,3560336	−50,9543289 +85,5656051 +34,6112762

Examinînd rezultatele obținute prin metodele de calcul aplicate, se constată o diferență foarte mare între ele. Astfel, ritmul mediu calculat prin metoda geometrică marchează valoarea 1,892, prin media parabolică — 2,009 și 1,127, prin metoda trendului exponențial. Deoarece majoritatea coeficienților individuali ai ritmului sînt mai apropiați de 1,127 decît de valorile obținute prin aplicarea celorlalte două metode, trendul exponențial se impune ca cel mai bun instrument de caracterizare a dezvoltării fenomenelor studiate. Reprezentarea grafică (fig. 7.7.) ne înfățișează, intuitiv, aceste deosebiri.

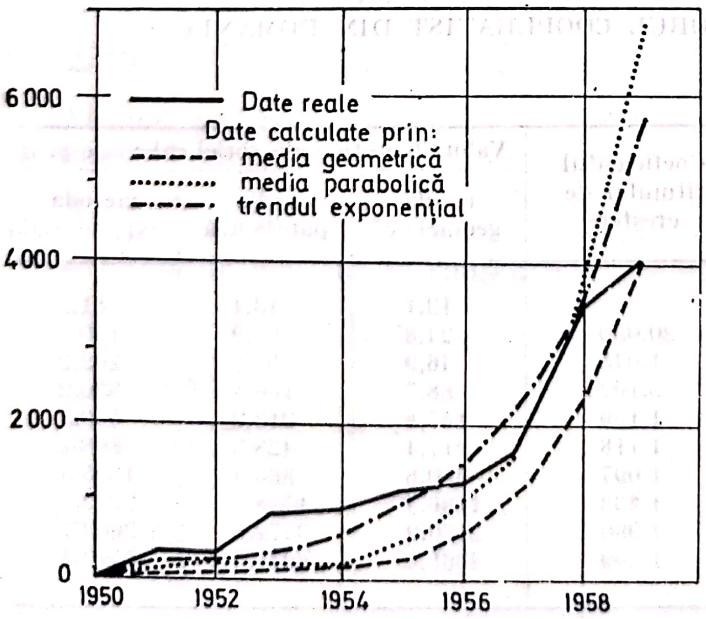


Fig. 7.7 Suprafața arabilă a sectorului cooperatist din România, la începutul anilor 1950—1959. După I. Marinescu [73].

În analiza ce se efectuează cu ocazia unor asemenea calcule statistice, se recomandă să se menționeze coeficienții cu valori extreme, de pildă, în cazul concret, 20,03 corespunzător anilor 1950—1951 și 3,102 — anilor 1952—1953.

În concluzie, cu privire la metodele de calcul al ritmului mediu de creștere, se poate spune că, deoarece acest calcul înseamnă aproximarea evoluției reale a seriei dinamice printr-o schemă teoretică oarecare, printr-un model teoretic, determinarea tendinței centrale a fenomenului studiat prin metoda celor mai mici pătrate ne oferă cel mai bun criteriu de aproximare. Analiza calitativă prealabilă a fenomenului cercetat, pe baza legilor obiective ce guvernează desfășurarea lui, va indica tipul de curbă care aproximează mai bine tendința centrală; după cum s-a mai spus, această metodă poate reflecta realitatea numai în condițiile unei continuități, a unei anumite stabilități în dezvoltarea fenomenului.

Determinarea ritmului mediu de creștere la fenomenele complexe. Prezentarea metodologiei calculului ritmului mediu de creștere, astfel cum a fost făcută mai înainte, a pornit de la ipoteza că indicatorii urmăriți în configurația lor de serie cronologică au o structură simplă și că deci ritmul dezvoltării lor poate fi considerat în mod global. În realitate, fenomenele social-economice se înfățișează ca sisteme complexe, cu structuri determinate de interdependența unor factori, adesea multipli și variați. În asemenea condiții se pune problema corelării ritmului mediu al fenomenului complex cu ritmurile medii ale componentelor factoriale.

Ne găsim deci în fața unor *sisteme de indicatori*; de pildă, producția este tratată statistic ca produs între productivitatea muncii și numărul mediu scriptic al angajaților, respectiv al muncitorilor; parcursul vehiculelor — ca produs între timpul mediu de mișcare pe parcurs al unui vehicul, viteza medie de exploatare și numărul mediu al vehiculelor puse zilnic în exploatare; volumul total al desfacerilor de mărfuri cuprinde, în realitate, însumarea desfacerilor de mărfuri industriale, alimentare, de alimentație publică ș.a.m.d.

Perfecționarea metodologiei calculului statistic al ritmului mediu de creștere a avut în vedere găsirea unor procedee cît mai sensibile la oscilațiile mari ale seriilor cronologice, fără calcule prea complicate, precum și a acelor metode care să poată da expresie structurii indicatorului complex studiat. În acest scop s-a făcut apel la funcțiile compuse, care matematic pot exprima relațiile de sumă sau produs existente între componentele sistemelor de indicatori, și anume:

$$x = f(u,\ v,\ z) = u(t) + v(t) + z(t) \tag{7.1}$$

$$x = f(u,\ v,\ z) = u(t) \cdot v(t) \cdot z(t). \tag{7.2}$$

Potrivit celor expuse pînă aici, calculul ritmului mediu de creștere a indicatorilor din aceste două formule presupune cunoașterea valorii termenului inițial și al celui final din serie; atunci cînd aplicăm metoda mediei geometrice,

Scanned with OKEN Scanner

a valorii tuturor termenilor, se utilizează media parabolică și trendul exponențial (prin aplicarea metodei celor mai mici pătrate). În multe cazuri aceste condiții nu sînt date, dar, utilizînd unele proprietăți ale ritmurilor medii, cercetarea statistică poate găsi soluții pentru rezolvarea problemei de calcul.

În cazul relației (7.1) de mai sus, pentru momentul t putem scrie:

$$x_0 \bar{\theta}_x^t = u\bar{\theta}_u^t + v\bar{\theta}_v^t + z\bar{\theta}_z^t; \tag{7.3}$$

deci

$$\bar{\theta}_x = \sqrt[t]{\frac{u_0\bar{\theta}_u^t + v\bar{\theta}_v^t + z_0\bar{\theta}_z^t}{u_0 + v_0 + z_0}} \tag{7.4}$$

și, întrucît

$$\bar{\theta}_u^t = \frac{u_n}{u_0}; \quad \bar{\theta}_v^t = \frac{v_n}{v_0}; \quad \bar{\theta}_z^t = \frac{z_n}{z_0},$$

formula devine:

$$\bar{\theta}_x = \sqrt[t]{\frac{u_n + v_n + z_n}{u_0 + v_0 + z_0}} = \sqrt[t]{\frac{x_n}{x_0}}, \tag{7.5}$$

relațiile (7.4) și (7.5) verificîndu-se și în cazul utilizării trendului exponențial la determinarea ritmului mediu de creștere a indicatorilor u, v și z, cînd se folosesc ca ponderi valorile inițiale ale acestora, din seria ajustată prin metoda celor mai mici pătrate.

Pentru calculul ritmului mediu al unui indicator component al sistemului, de pildă $\bar{\theta}_z$, cunoscîndu-se $\bar{\theta}_x$, $\bar{\theta}_u$ și $\bar{\theta}_v$, se recurge la formula

$$\bar{\theta}_z = \sqrt[t]{\frac{x_0\bar{\theta}_x^t - (u_0\bar{\theta}_u^t + v_0\bar{\theta}_v^t)}{z_0}} \tag{7.6}$$

și apoi, simplificînd prin înlocuire,

$$\bar{\theta}_z = \sqrt[t]{\frac{z_n - (u_n + v_n)}{z_0}}. \tag{7.7}$$

În cazul relației (7.2) procedeul de calcul este similar, la momentul t ea devenind:

$$x_0\bar{\theta}_x^t = u_0\bar{\theta}_u^t \cdot v_0\,\bar{\theta}_v^t \cdot z_0\,\bar{\theta}_z^t, \tag{7.8}$$

care poate fi scrisă și sub forma:

$$\bar{\theta}_x^t = \bar{\theta}_u^t \cdot \bar{\theta}_v^t \cdot \bar{\theta}_z^t, \tag{7.9}$$

de unde

$$\bar{\theta} = \sqrt[t]{\bar{\theta}_u^t \cdot \bar{\theta}_v^t \cdot \bar{\theta}_z^t} = \theta_u \cdot \theta_v \cdot \theta_z. \tag{7.10}$$

Scanned with OKEN Scanner

Pentru demonstrare, exemplificăm cu sistemul de indicatori din tabelul 7.7.

Tabelul 7.7

PRINCIPALII INDICATORI AI ACTIVITĂȚII UNUI PARC DE MIJLOACE DE TRANSPORT ÎN COMUN

date convenționale)

Perioada (t)	Parcursul vehiculelor veh/km (x)	Viteza medie de exploatare km/oră (u)	Timp mediu de mișcare pe parcurs al unui vehicu , în ore (v)	Nr. mediu de vagoane zilnic în circulație (z)
1	9 625,0	12,5	11,0	70
2	10 744,8	12,1	11,1	80
3	11 328,0	11,8	12,0	80
4	12 825,0	11,4	12,5	90
5	14 560,0	11,2	13,0	100

Analiza calitativă prealabilă învederează existența următoarelor relații între indicatorii factoriali ai fenomenului complex:

$$S = f(v, z) = v(t)z(t),$$

în care S înseamnă vehicule-ore pe zi;

$$x = \varphi(u, v, z) = u(t)v(t)z(t) = u(t)S(t)$$

$$\bar{\theta}_u = 0{,}9739; \quad \theta_v = 1{,}0423; \quad \theta_z = 1{,}0933.$$

Aplicînd relația

$$\bar{\theta}_x = \bar{\theta}_u \cdot \bar{\theta}_v \cdot \bar{\theta}_z = 0{,}9739 \cdot 1{,}0423 \cdot 1{,}0933 = 1{,}109.$$

Ritmul mediu pentru indicatorul S, conform uneia din relațiile de mai sus, se poate calcula cunoscînd ritmurile medii ale indicatorilor factoriali v și z. De asemenea, se pot calcula și ritmurile medii pentru relațiile succesive dintre indicatori, de pildă:

$$S = \varphi(v, z) = v(t)z(t)$$

$$x = f(u, v, z) = u(t) \cdot v(t) \cdot z(t) = S(t)u(t).$$

După cum se vede, calculul este mai ușor în cazul sistemelor de forma relației (7.2), întrucît se utilizează ca elemente numai ritmurile medii ale indicatorilor factoriali.

Aplicație a ritmului mediu de creștere în comparații internaționale. Statistica internațională se dezvoltă, în primul rînd, ca o cercetare comparativă complexă a diferitelor fenomene, pe țări, pe grupe de țări sau pe zone cu caracteristici comune. În acest context, ritmul de creștere este folosit, de pildă, ca un indicator al măsurării decalajului în ce privește dezvoltarea producției unui anumit produs x, dintre țările A și B. Dacă urmărim cunoașterea mo-

Scanned with OKEN Scanner

mentului în care ţara A va atinge nivelul producţiei respective din ţara B, vom proceda în felul următor [65]:

Notînd cu x_0^A şi x_0^B producţia din anul de bază şi estimînd ritmul de creştere θ al producţiei în cele două ţări, se va obţine egalitatea

$$x_0^A \, \bar{\theta}_A^t = x_0^B \, \bar{\theta}_B^t$$

atunci cînd se va atinge acelaşi nivel de producţie în ambele ţări.

Prin logaritmare, ecuaţia devine

$$\log x_0^A + t \log \bar{\theta}_A = \log x_0^B + t,$$

care se poate scrie:

$$t \log \bar{\theta}_A - t \log_B = \log x_0^B - \log x_0^A = t(\log \bar{\theta}_A - \log \bar{\theta}_B) = \log x_0^B - \log x_0^A,$$

iar

$$t = \frac{\log x_0^B - \log x_0^A}{\log \bar{\theta}_A - \log \bar{\theta}_B}.$$

Exemplificînd cu date numerice:

Producţia în tone, 1975		θ
Ţara A	300	1,15
Ţara B	500	1,10

vom obţine:

$$t = \frac{\log 500 - \log 300}{\log 1{,}15 - \log 1{,}10} = \frac{2{,}69897 - 2{,}47712}{0{,}06070 - 0{,}04139} = 11{,}5 \text{ ani},$$

adică timpul în care ambele ţări vor atinge un nivel egal al producţiei.

Invers, dacă se caută ritmul necesar pentru a înlătura decalajul, dorindu-se ca ţara A să ajungă la acelaşi nivel cu ţara B, calculul se va desfăşura pe baza formulei:

$$t \log \bar{\theta}_A = \log x_0^B - \log x_0^A + t \log \bar{\theta}_B$$

şi se obţine $\bar{\theta}_A = 1{,}15$.

Printre aplicaţiile cele mai cunoscute ale ritmului mediu de creştere se notează însă utilizarea lui la operaţiile de interpolare şi extrapolare.

De multe ori, problema se pune în alţi termeni. De pildă, care este numărul de ani peste care valoarea unui fenomen devine dublă, în condiţiile unui ritm mediu de creştere dat? Aşa este cazul populaţiei mondiale al cărei ritm mediu anual de creştere este în prezent de 2 la sută anual. Problema se rezolvă astfel:

$$(1 + r)^x = 2,$$

în care:

x este numărul de ani peste care se dublează valoarea indicatorilor;
2 — dublarea
r — ritmul mediu dat.

Scanned with OKEN Scanner

Prin logaritmare, vom avea

$$x \log (1 + r) = \log 2$$

$$x = \frac{\log 2}{\log (1 + r)}.$$

În exemplul nostru, în care $r = 2\%$, adică $(1 + r) = 1{,}02$

$$x = \frac{\log 2}{\log 1{,}02} = \frac{0{,}30103}{0{,}00860} = 35 \text{ ani.}$$

Venitul național al țării noastre urmează să crească în perioada 1976—1980 cu un ritm mediu anual de 11%. În aceste condiții, el se poate dubla astfel:

$$x = \frac{\log 2}{\log 1{,}11} = \frac{0{,}30103}{0{,}04532} = 6{,}6 \text{ ani.}$$

De obicei, se dau tabele cu valori pentru situațiile cele mai frecvente: $r = 0{,}5;\ 1{,}0;\ 1{,}5;\ 2{,}0$ etc.

Interpolare și extrapolare. Derivat din termenul latin *interpolare* (a reînoi) și cunoscut deja indienilor, iar în secolul al 13-lea și arabilor, procedeul *interpolării* reprezintă o metodă algebrică de introducere (completare) într-o serie de valori a unuia sau mai multor termeni, determinați direct prin calcul numeric. Cu alte cuvinte, dacă valorile unei funcții de o singură variabilă x — să notăm u_x — au fost tabelate pentru valori echidistante ale variabilei x, $x + h$, $x + 2h$ ș.a.m.d., operația de interpolare constă în a afla valori ale funcției corespunzătoare unei valori intermediare a variabilei. În mod obișnuit se face apel la interpolare în următoarele cazuri: 1) pentru intercalarea de noi valori intermediare între datele empirice ale seriei; 2) cînd seria se prezintă sub formă de grupe de date (de exemplu, grupe de vîrstă într-o serie demografică), pentru a obține elementele grupelor; 3) în scopul simplificării unei curbe, ale cărei variații au o formă prea complicată pentru analiză [134].

Metodă matematică de aproximare, interpolarea dă bune rezultate în științele naturii, precum fizica sau astronomia, dar eficiența ei în metodologia seriilor cronologice este condiționată de o variație absolut lină a acestora, ceea ce se poate obține numai printr-o netezire[3] deliberată, pe baza construirii unui algoritm adecvat. Problema care trebuie să fie soluționată prin interpolare constă în determinarea unei astfel de curbe, încît să fie cea mai apropiată de ansamblul de puncte dintre extremitățile ordonatelor care conduc la historiograma seriei.

Spre deosebire de interpolare, prin *extrapolare* se caută să se obțină acele valori care prelungesc seria dincolo de limitele termenilor dați ai seriei cronologice. Extrapolarea este folosită curent pentru previziuni și prognoze.

În operația de interpolare se utilizează polinoamele, diferite funcții (exponențială, logaritmică, sinusoidală), găsindu-și aplicarea și aici metoda celor mai mici pătrate, mai rar metoda celor mai mici distanțe și alte procedee. Întrucît cel mai des folosite sînt polinoamele, prezentăm un procedeu de calcul.

[3] În engleză, *smoothing* înseamnă ansamblu de procedee pentru construirea unei curbe line într-o serie statistică, ale cărei date empirice se succed prin vîrfuri ascuțite sau căderi neașteptate; termen echivalent cu cel de *ajustare*. Termenul adoptat în statistica matematică este de *netezire*.

Problema care se pune este de a găsi un polinom de grad n, cunoscînd valorile pe care le ia pentru $n+1$ valori date ale argumentului. Aceasta ne conduce la *formula de interpolare a lui Lagrange*[4], şi anume:

Considerînd $x_0, x_1, x_2, \ldots, x_n$, valorile argumentului $y_0 = f(x_0)$; $y_1 = f(x_1)$; $y_2 = f(x_2), \ldots, y_n = f(x_n)$, vrem să aflăm pe $y = f(x)$. Trebuie să se găsească întîi un polinom $f_0(x)$, care să întrunească condiţiile: $f_0(x) = y_0$, pentru $x = x_0$; $f_0(x) = 0$, pentru $x = x_1, x_2, \ldots, x_n$.

Întrucît $f_0(x_1) = 0$, $f_0(x_2) = 0, \ldots, f_0(x_n) = 0$, $f(x)$ fiind un polinom, rezultă că $f(x)$ este divizibil prin $(x - x_1)$, $(x - x_2), \ldots, (x - x_n)$, adică produsul $(x - x_1)\ (x - x_2), \ldots, (x - x_n)$.

Luînd polinomul de gradul cel mai mic, adică de gradul n, vom avea $f_0(x) = k(x - x_1)(x - x_2) \ldots (x - x_n)$, urmînd a fi determinat coeficientul k.

Întrucît $f_0(x_0) = y_0$, deci $f_0(x_0) = k(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)$

$$k = \frac{y_0}{(x_0 - x_1)(x_0 - x_2) \ldots (x_{0-n})},$$

iar funcţia căutată în cazul particular:

$$f(x) = y_0 \frac{(x - x_1)(x - x_2) \ldots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)}.$$

Pentru funcţiile $f_1(x), f_2(x), \ldots, f_n(x)$, egale cu $y_1, y_2, \ldots, y_n$, pentru x, respectiv egal cu $x_1, x_2, \ldots, x_n$ şi egale cu 0 pentru celelalte valori, vom avea

$$f_1(x) = y_1 \frac{(x - x_0)(x - x_2) \ldots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \ldots (x_1 - x_n)}$$

$$f_2(x) = y_2 \frac{(x - x_0)(x - x_1)\ (x - x_3) \ldots (x - x_n)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3) \ldots (x_2 - x_n)}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$f_n(x) = y_n \frac{(x - x_0)(x - x_1) \ldots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \ldots (x_n - x_{n-1})}.$$

Suma polinoamelor $F(x) = f_0(x) + f_1(x) + f_2(x) + \ldots + f_n(x)$ ne dă ca rezultat un polinom de gradul n, care pentru $x - x_0$ ia valoarea y_0, pentru $x - x_1$ — valoarea y_1, pentru ... $x = x_n$ — valoarea y_n şi, în general, pentru $x = x_i$ — valoarea y_i, deci

$$f_i(x_i) = y_i,$$

întrucît

$$F(x_i) = f_0(x_i) + f_1(x_i) + f_2(x_i) + \ldots + f_n(x_i).$$

$F(x)$ ne dă tocmai polinomul căutat $f(x)$. Polinomul de interpolare va avea deci forma:

$$f(x) = y_0 \frac{(x - x_1)(x - x_2) \ldots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)} + y_1 \frac{(x - x_0)(x - x_2) \ldots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \ldots (x_1 - x_n)} + \ldots$$

$$\ldots + y_n \frac{(x - x_1)(x - x_2) \ldots (x - x_{n-1})}{(x_n - x_1)(x_n - x_2) \ldots (x_n - x_{n-1})}$$

[4] LOUIS DE LAGRANGE (1736—1813), astronom şi matematician francez, a desfăşurat o activitate considerabilă, rămînînd cunoscut, între altele, prin contribuţia sa la calculul variaţional şi funcţiile derivate.

Scanned with OKEN Scanner

sau

$$f(x) = \sum_{i=0}^{n} y_i \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x_i - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

Pentru exemplificarea calculului se folosesc datele din tabelul 7.8. Se cere să stabilim care a fost producția totală de combustibili în anul 1968 (interpolare) și care va fi producția în anul 1978 (extrapolare).

Tabelul 7.8

PRODUCȚIA SURSELOR DE ENERGIE PRIMARĂ ÎN GRUPA DE ȚĂRI X, ÎN ANII 1965, 1970 și 1975

(în milioane tone echivalent cărbune)

Surse de energie primară	1965	1970	1975
Total, din care:	178,41	234,98	302,0
Huilă	99,91	121,38	138,9
Lignit	65,39	95,82	123,3
Combustibili lichizi	7,72	16,37	18,5
Gaze naturale	4,59	8,79	18,8
Hidroelectricitate	0,80	1,62	2,6

În operațiile de calcul pornim de la următoarele valori pentru x și y:

x	$x = 0$	$x_1 = 5$	$x_2 = 10$
y	178,41	243,98	302,0

Polinomul de interpolare va fi de gradul II; utilizînd cele trei valori cunoscute ale funcției:

$$y = y_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + y_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + y_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} =$$

$$= 178{,}41 \frac{(x - 5)(x - 10)}{50} + 234{,}98 \frac{x(x - 10)}{-25} + 302 \frac{x(x - 5)}{50} =$$

$$= 3{,}5682(x^2 - 15x + 50) - 9{,}7592(x^2 - 10x) + 6{,}04\ (x^2 - 5x) =$$

$$= -0{,}151x^2 + 13{,}869x + 178{,}41.$$

Producția din anul 1968 corespunde valorii nr. 3, deci valoarea funcției se calculează dînd lui x această valoare în polinomul de interpolare, adică:

$$y_{1968} = 0{,}151 \times 9 + 13{,}869 \times 3 + 178{,}41 = 218{,}658,$$

care este volumul producției totale de energie primară atins în 1968, stabilit prin interpolare.

Producția pe 1978 (previziune) se calculează făcînd pe $x = 13$, adică $1978 - 1965 = 13$, deci

$$y_{1978} = -0{,}151 \times 169 + 13{,}869 \times 13 + 178{,}41 = 333{,}188,$$

care este volumul producției totale previzionate prin operația de extrapolare.

Ritm și tendințe

Scanned with OKEN Scanner

O altă formulă clasică, utilizată în operaţiile de interpolare, este ecuaţia cunoscută sub denumirea de *regula* sau *formula lui Newton:*

$$u_x = u_0 + x\Delta_0^1 + \frac{x(x-1)}{1\cdot 2}\Delta_0^2 + \frac{x(x-1)(x-2)}{1\cdot 2\cdot 3}\Delta_0^3,$$

dedusă din tabelul ce poartă numele aceluiaşi fizician, astronom şi matematician englez celebru.[5].

Pentru utilizarea ritmului mediu de creştere în operaţiile de interpolare şi extrapolare, vom porni de la formulele deja cunoscute, dînd şi exemplele ce urmează.

Astfel, de pildă, dacă cunoaştem numărul populaţiei ţării noastre în anii 1948 (15872624 locuitori) şi 1956 (17489450 locuitori) şi vrem să stabilim cifra intermediară pentru anul 1954, vom calcula mai întîi ritmul mediu de creştere, adică:

$$\bar{\theta} = \sqrt[8]{\frac{17489450}{15872624}} = 1{,}0121 \text{ sau } 12{,}1\,\%,$$

astfel că ritmul mediu al sporului va fi $r = 0{,}0121$.

Pentru calculul numărului populaţiei în 1954, vom folosi formula:

$P_{1954} = P_{1948}\cdot\bar{\theta}^t = 15872\cdot 1{,}0121^6 = 17048000$ locuitori, cifră care nu diferă prea mult de cea din statistica oficială (17040000)[6].

Exemplificînd aplicarea ritmului de creştere în operaţia de extrapolare cu un calcul demografic similar, în care, cunoscîndu-se ritmul mediu de creştere, de 1,0098 şi numărul populaţiei din anul 1960 (18403434 locuitori), se cere să se stabilească numărul ei pentru 1964, vom avea:

$$P_{1964} = P_{1960}\cdot\bar{\theta}^t = P_{1960}\cdot\bar{\theta}^4 = 18403434\cdot 1{,}0098^4$$

$$\log 1{,}0098^4 = 4\log 1{,}0098 = 0{,}01692.$$

Antilogaritmul fiind

$$1{,}0098^4 = 1{,}0399,$$

$$P_{1964} = 18403434 \times 1{,}0399 = 19137731.$$

Confruntînd ultima cifră cu populaţia evidenţiată pentru anul 1964 în anuarul statistic, adică 18927081 locuitori, constatăm o diferenţă de aproape 211000 locuitori. Pentru a găsi explicaţia acestei diferenţe apreciabile urmărim evoluţia seriilor cronologice ale mişcării naturale a populaţiei, care este ilustrată sugestiv în graficul din fig. 7.8. Evoluţia oscilatorie a curbei deceselor, dar mai ales cea constant regresivă a născuţilor vii în deceniul 1955—1964 ne conduce la concluzia că în calculul de mai sus s-a operat cu un ritm mediu de creştere inexact, analiza calitativă prealabilă a desfăşurării fenomenului nefiind suficient de aprofundată.

În concluzie, se poate arăta că utilizarea interpolării şi extrapolării la seriile cronologice necesită multă circumspecţie şi răspundere. Mult criticată, extrapolarea este totuşi frecvent folosită şi continuă să rămînă o metodă fundamentală de cercetare ştiinţifică a viitorului. Criticile care i se aduc nu vizează

[5] Deducerea şi aplicaţii ale formulei în vol. [134].
[6] *Breviar statistic al R.P.R. 1961.*

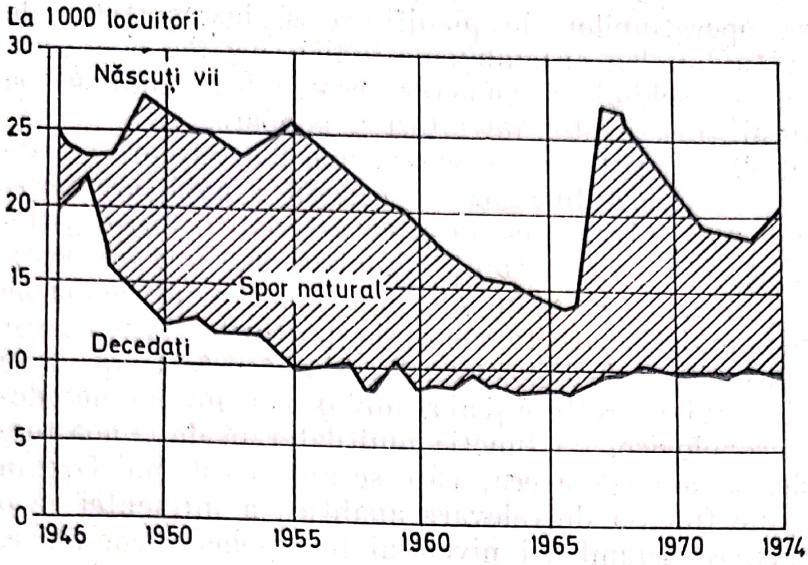


Fig. 7.8 Mișcarea naturală a populației României, în anii 1966—1974.
Sursa: Anuarul statistic al R. S. România 1975.

atît metoda în sine, cît aplicarea ei. „Problema cuantificării factorilor sociali, de pildă, sau problema posibilității izolării unui ansamblu de elemente într-un context dat, nu sînt probleme specifice ale extrapolării; erorile în astfel de etape ale studiului *nu* sînt erori ale metodei de extrapolare, ci erori *de modelare* cantitativă"[7]. În domeniul proiectărilor demografice, s-a constatat că metoda progresiei geometrice prezintă un mare risc și, în general, metodele matematice de proiectare a numărului populației au în prezent o aplicare mai redusă. Se ia tot mai mult în considerare influența factorilor social-economici și se afirmă pe primul plan utilizarea componentelor (a metodei analitice), caracterizată prin poiectarea subpopulațiilor (pe efective de ani sau grupe de ani, vîrstă, sexe) ca bază a proiectării numărului populației [30].

7. 3 METODA INDICILOR STATISTICI

7.3.1 Caracteristici, retrospectivă, clasificări

Alături de mărimile medii, metodologia statisticii acordă de mai mult timp o importanță din ce în ce mai mare *indicilor*. Prin ei, limbajul statistic reflectă cu multă expresivitate schimbările din viața socială, ceea ce l-a făcut pe statisticianul Helmut Swoboda să afirme: „Indicele este degetul arătător al economiei, indicatorul progresului și al insuccesului... El devine caracteristic pentru o întreagă situație" [120].

Cu ajutorul indicilor statistici se caracterizează evoluția fenomenelor în timp, se compară manifestările lor în spațiu, se măsoară influența ce o exercită diferiți factori asupra nivelului și structurii fenomenelor complexe. Ei exprimă caracteristicile dezvoltării vieții social-economice sub cele mai diverse aspecte: producție și investiții, productivitatea muncii și nivel de trai, venituri și retribuirea muncii, circulația bunurilor și persoanelor, circulația monetară, activitatea financiară și de credit etc. În economia socialistă, indicii

[7] M. BOTEZ, *Metoda extrapolării* [30].

Scanned with OKEN Scanner

statistici constituie baza operaţiunilor de planificare şi instrumentul de verificare a îndeplinirii planului. Un ansamblu de indici specifici oglindesc, în timp şi spaţiu, relaţiile de schimb şi cooperare economică, ştiinţifică şi tehnică, cu caracter cultural, turistic etc., între ţări şi popoare.

Se vede, astfel, că indicele, statistic are o extensiune foarte mare şi este totdeauna o medie. El caracterizează dinamica unei medii, a unei caracteristici într-o colectivitate omogenă, nivelul mediu al caracteristicii într-o astfel de colectivitate, dar cu o structură în continuă schimbare sau într-o colectivitate eterogenă, precum şi schimbarea volumului unei colectivităţi omogene sau eterogene. De asemenea, indicele caracterizează şi raporturile dintre aceste elemente şi aspecte, iar în calitate de indicator al schimbării în timp, el se identifică cu indicatorul ritmului de creştere şi al ritmului sporului din metodologia statistică a seriilor cronologice. La funcţia multilaterală de cunoaştere ce o îndeplineşte indicele, se adaogă aceea, care se aplică tot mai larg în practica statistică, şi anume funcţia de relevare analitică a influenţei ce o exercită diferiţi factori asupra schimbării nivelului fenomenelor complexe. Funcţia cea mai importantă a indicelui statistic este însă aceea de măsurare a schimbării în timp a unui fenomen oarecare, fiind legat deci de seriile cronologice. Cu asemenea calităţi indicii joacă un rol important în luarea deciziilor de politică economică.

Sintetizînd caracteristicile relevate într-o definire cît mai generală, putem arăta că indicele statistic este un indicator, o categorie statistică, exprimată într-o mărime sintetică, utilizată pentru a reda variaţia relativă între două stări, una de actualitate, cealaltă de referinţă, a unui fenomen, în timp, în spaţiu sau în procesul de realizare a planului.

În sens matematic, indicele statistic este un număr pur, care măsoară variaţiile relative ale unei mărimi ce reprezintă fie o caracteristică statistică x, fie frecvenţa f, fie agregate sau complexe de tipul xf şi Σxf; el se construieşte pe bază de valori absolute, relative sau medii. În mod simplu, variaţia mărimii se va exprima sub forma raportului

$$I_{1/0} = \frac{M_1}{M_0}, \quad \text{respectiv} \quad I_{1/0} = \frac{M_1}{M_0} \cdot 100,$$

în care M_0 şi M_1 reprezintă măsura mărimii, corespunzătoare celor două fracţiuni de timp sau elementelor de spaţiu ce se compară — cea de referinţă t_0 şi cea actuală t_1. Dar fenomenele social-economice nu ne oferă, de obicei, mărimi simple, ci ansambluri, colectivităţi (mulţimi) de elemente şi pentru a da expresie întregului ansamblu recurgem la construirea, prin abstracţiune, a indicilor de grup (sintetici, compuşi). În indicele valorii unei mărfi, de exemplu, este inclusă variaţia cantităţii şi a preţului ei; indicele producţiei industriale sintetizează producţia diferitelor ramuri, a diferitelor sortimente etc. În literatura statistică se face uneori distincţie între *indicii sintetici*, în care se consideră a fi întruniţi, printr-un calcul de medie, indici particulari, reali, de aceeaşi natură (de pildă, preţurile de gros) şi *indici compuşi*, care grupează într-o singură expresie numerică indici cu caracter diferit, dar care se pot încadra, prin caracteristici comune, într-un anumit fenomen (de exemplu, producţia unei întreprinderi industriale).

Construirea indicilor de grup a necesitat, în cursul timpului, rezolvarea a numeroase probleme teoretice şi metodologice, care nici astăzi nu sînt definitiv soluţionate. În statistica ţărilor capitaliste au apărut inconveniente,

legate de utilizarea în practică a indicilor, şi anume acolo unde se cereau reguli rigide (de pildă, *indexarea* salariilor sau a chiriilor, contracte *indexate* etc.). O perioadă îndelungată, de peste două secole, un mare număr de economişti şi matematicieni şi-au consacrat eforturile pentru găsirea de procedee adecvate pentru construirea indicilor de grup, care să corespundă conţinutului real al fenomenelor. A fost elaborat un număr mare de formule, dintre care economistul american Irving Fisher a analizat 134, la care a adăugat şi formula sa [8].

În practica statistică, indicii au apărut în formă de indici *simpli* (individuali), în scopul comparării preţurilor, la epoci diferite, la bunuri de consum de strictă necesitate: grîu, carne, pîine sau cărbune. Se menţionează o asemenea încercare încă în anul 1707, în Anglia, ea fiind efectuată de episcopul Fleetwood, în *Chronicon Preciosum* [120]. Despre construirea unui indice *de grup* se poate vorbi de la anul 1738, cînd francezul Dutot s-a ocupat de o problemă asemănătoare, folosind o listă de mărfuri mai extinsă şi calculînd indicele după următorul procedeu [120]:

$$I = \frac{\dfrac{\text{Preţuri noi}}{\text{Preţuri vechi}}\ \dfrac{15}{10} + \dfrac{14}{8} + \dfrac{80}{60}}{\text{N-rul mărfurilor}} = \frac{\dfrac{18}{12} + \dfrac{21}{12} + \dfrac{16}{12}}{3} = \frac{55}{36} = 1{,}525.$$

Astfel, într-o formulă generală, dacă notăm cu p_1, p_2, ..., p_n, preţurile diverselor mărfuri din perioada de bază şi cu P_1, P_2, ..., P_n preţurile aceloraşi mărfuri din perioada curentă, indicele va fi:

$$I = \frac{p_1 + p_2 + \ldots + p_n}{P_1 + P_2 + \ldots + P_n} = \frac{p}{P}.$$

Întrucît fiecare termen poate fi împărţit la n, formula capătă aspectul unui cît (raport) de medii. El poate fi prezentat şi astfel:

$$I = \frac{\frac{1}{n}\Sigma p}{\frac{1}{n}\Sigma P}$$

şi mai poartă numele de *indice Bradstreet*, fiind utilizat în Anglia începînd din 1895, pentru a defini indicele de preţuri publicat de revista „Bradstreet". La utilizarea acestui indice s-a constatat de îndată o dificultate, şi anume s-a născut problema dacă pentru unitatea de greutate la care se referă preţul este de preferat cota respectivă în kilograme sau în tone, deoarece făcîndu-se calculul pe baza preţurilor unor unităţi de măsură diferite, s-ar obţine rezultate diferite. De asemenea, la construirea indicilor menţionaţi pînă acum nu erau luate în considerare deprinderile (obiceiurile) în materie de consum şi schimbările lor.

Un indice mai cunoscut a fost cel propus în 1764 de contele Gianrinaldo Carli-Rubbi [9], istoric, astronom şi economist italian, într-un studiu cu caracter

[8] IR. FISHER, *The making of index-numbers*, New York, 1922.

[9] Lucrarea sa, *Delle monete e dell'istituzione delle zecche d'Italia dell'antico e presente sistema d'esse: e del loro intrinseco valore, e raporto con la presente moneta. Dalla decadenza dell'impero sino al secolo XVII. Per utile della Publiche, e delle Private Ragioni. Dissertazioni*, apărută la Mantova în 1754, anticipează necesitatea calculării indicelui de variaţie a preţurilor.

Scanned with OKEN Scanner

monetar, în care a cercetat deprecierea monetară din Italia în perioada dintre 1500 și 1750. Carli a calculat, pentru fiecare produs, un indice elementar al prețurilor și indicii astfel obținuți au fost grupați cu ajutorul unei medii aritmetice. Formula apare astfel:

$$I = \frac{1}{n}\left(\frac{p_1}{P_1} + \frac{p_2}{P_2} + \ldots + \frac{p_n}{P_n}\right) = \frac{1}{n}\,\frac{p}{P}.$$

Aici avem de-a face cu o medie a cîturilor, în care unitatea de cantitate a fiecărei mărfi nu mai are influență în determinarea prețului. Nici Carli, cum se vede, n-a folosit un sistem de ponderare.

Statistica generală a Franței a utilizat, vreme îndelungată, această formulă (care, în fond, este o medie aritmetică de prețuri relative) la calcularea indicelui prețurilor de gros pentru 45 mărfuri, formula devenind în acest caz:

$$I = \frac{1}{45}\left(\frac{p_1}{P_1} + \frac{p_2}{P_2} + \ldots + \frac{p_n}{P_n}\right).$$

S-a observat însă că dacă se stabilesc prețurile corespunzătoare cantităților, în așa fel încît $P_1 = P_2 = \ldots P_n$, cele două formule dau un rezultat identic, pentru că:

$$\frac{\Sigma p}{\Sigma P} = \frac{1}{np}\,\Sigma P = \frac{1}{n}\sum\frac{p}{P}.$$

A trebuit să se recunoască apoi că, în fapt, se operează cu o valoare medie a articolelor la anumite cantități, și nu cu prețul lor. În aceste condiții posibilitățile de utilizare a primei formule s-au îngustat. De asemenea, dacă după prima metodă unitatea de măsură acordă o importanță diferită anumitor produse, a doua metodă le reduce la o uniformitate ce nu se putea accepta; prețul pîinii nu avea mai mare importanță decît, de exemplu, cel al piperului și variația prețului la diversele articole nu avea nici o influență asupra indicelui general. Astfel a apărut necesitatea atribuirii unui „loc preponderent" anumitor articole și aceasta a condus la ponderarea prețurilor, ținîndu-se seama de importanța relativă a diverselor mărfuri, prin afectarea unui anumit coeficient. Se pare că cel dintîi care a recurs la această soluție a fost Arthur Young. Introducînd un coeficient k, prima formulă a devenit:

$$I = \frac{\Sigma pk}{\Sigma Pk}.$$

Acest coeficient nu era însă corect determinat. Practic, trebuia să se ia în considerare diverse cantități pentru diverse mărfuri. Astfel, în cursul timpului, a apărut în diferite țări o sumedenie de modalități de calculare a indicelui variației prețurilor, în funcție de diverse criterii și scopuri [10].

[10] Din necesitatea de a măsura și caracteriza variația curentă a prețurilor mărfurilor și a cursurilor valorilor imobiliare (acțiuni și obligațiuni) cotate la bursă, apoi în mod global de a cunoaște conjunctura pieței, au apărut în țările capitaliste diverși indici, începînd cu mijlocul secolului al XIX-lea, unii din ei fiind calculați și în prezent. Indicele construit de *William Newmarch*, pentru a exprima variația prețurilor și publicat în 1859 în revista „Journal of the Statistical Society", a fost numit de el *Index-Number*, denumire care a rămas caracteristică în țările de limbă engleză, fiind preluată apoi și de alte țări în traducerea respectivă (franceză *nombre indice*, germană *Indexziffer*) și sinonimă cu termenul românesc de *indice*. Pentru urmărirea valorilor industriale se publică și azi, în S.U.A., indicele *Dow Jones*, calculat în 1889 de ziaristul

Complexitatea fenomenelor social-economice a căror schimbare constituie domeniul de aplicare a funcției indicilor, ca și problematica exprimării lor în termeni matematici determină o mare varietate a indicilor și a condus la o tipologie care se structurează după mai multe criterii, definite în felul următor:

a) *sfera de cuprindere*, în funcție de care indicii se împart în: indici elementari (individuali) și indici sintetici (de grup);

b) *perioada de referință*, după care diferențiem indicii cu bază fixă de indicii în lanț;

c) *procedeul de calcul*, potrivit căruia se construiesc indici de grup agregați și indici de grup medii armonici sau aritmetici ai indicilor individuali;

d) *structura colectivității*, care conduce la elaborarea de indici cu structură fixă, indici cu structură variabilă și indici de structură sau ai schimbărilor structurale;

e) *natura fenomenului*, a cărui variație își găsește expresia în indice, determină clasificarea în indici valorici, ai volumului fizic, ai prețurilor, ai productivității muncii, ai retribuirii medii etc.

La aprofundarea și dezvoltarea teoriei și metodologiei indicilor au adus o contribuție importantă statisticienii din țările socialiste. În ce privește clasificarea, s-a remarcat statisticianul sovietic *V.N. Peregudov* [11], care, plecînd de la criteriul funcției cognitive a indicilor, a stabilit următoarele patru grupe tipice: 1. indici care caracterizează cantitativ dezvoltarea fenomenului complex; 2. indici ai unor elemente izolate ale fenomenelor complexe; 3. indici de caracterizare a dinamicii medii a unui singur element (indici cu structură variabilă); 4. indici care caracterizează deplasările structurale.

Henry Charles Dow și ameliorat apoi de expertul financiar Jones. Institutul Harvard din S.U.A. a publicat, începînd din 1919, așa-numitul *barometru Harvard*, pe trei grupe — efecte, mărfuri, finanțe — caracterizînd conjunctura piețelor respective pentru perioada 1903—1914; după primul război mondial, acest indice s-a discreditat. În Franța, Institutul Național de Statistică și Studii Economice (I.N.S.E.E.) utilizează diverși indici pentru variația prețurilor de gros și de detail, denumiți după numărul de articole ce intră în componența lor. Astfel, *indicele celor 45 de mărfuri*, construit în 1911, sau *i. celor 319 mărfuri* (54 alimentare, 58 energetice, 207 industriale) se referă la prețurile de gros; pentru oglindirea costului vieții, cei mai cunoscuți indici ai prețurilor de detail sînt *indicele celor 13*, respectiv *al celor 34 articole de menaj* și *i. celor 179, 213*, respectiv *250 de articole*. În România antebelică, ziarul „Argus" publica un indice cu baza 1916, pentru *39 articole* (21 alimentare, 9 de îmbrăcăminte, 9 diverse), calculat după metoda *Viator*, iar fostul Institut central de statistică calcula cîte un *indice general* și *parțial* al prețurilor de detail, în București (baza 1929 = 100).

În principalele țări capitaliste se publică așa-numiții *indici internaționali de prețuri*, care de fapt reprezintă mișcarea prețurilor unor materii prime, produse de bază și semifabricate pe piețele capitaliste reprezentative (cu volum mare de tranzacții), folosind cotațiile de bursă, prețuri de la licitații sau de vînzare-cumpărare din contracte ș.a. Cei mai cunoscuți sînt indicii *Reuter* (*Commodity-Index*, calculat în Anglia din 1933), *Moody* (*Index of staple commodities*, elaborat în S.U.A. și avînd baza 31 decembrie 1931), *Volkswirt* și *H.W.W.A.* (*Hamburgisches-Weltwirtschaftliches-Archiv*), calculați în R. F. Germania din 1947 și respectiv 1954. Din același an 1954 calculează Biroul statistic al O.N.U. un indice pentru materii prime — *Index of Primary Commodities*, publicat trimestrial în „Monthly Bulletin of Statistics" și cuprinzînd 40 de mărfuri cu peste 250 serii de prețuri. Pentru circulația mărfurilor în comerțul exterior și pe plan mondial se folosesc *indicii raportului de schimb* (*Terms of trade*) și *raportului de volum*, iar pentru variațiile cheltuielilor de navlu în transporturile maritime, *indicii navlurilor*.

[11] V. N. PEREGUDOV, schema publicată în *Ucionîie zapiski po statistike*, vol. I, Moscova, 1955.

Scanned with OKEN Scanner

Tipologia complexă enunţată schematic invederează problematica construirii indicilor sintetici, care reclamă, de la început, precizarea fenomenului complex de studiat, cerînd un examen minuţios al scopului urmărit prin elaborarea indicelui, cuprinderea, într-o formulă adecvată, a elementelor comensurabile ce se asamblează în componenţa acestuia, alegerea unor elemente de ponderare în cadrul indicelui sintetic, stabilirea perioadei de timp pentru comparare (referinţă), care să corespundă caracteristicilor şi evoluţiei fenomenului. Validitatea unui indice, din punct de vedere formal, este condiţionată de calităţile de *reversibilitate* şi *transferabilitate*, asupra cărora ne vom opri în cele ce urmează.

7.3.2 Tehnica de calcul al indicilor

În contrast cu simplitatea schemei matematice a oricărui indice individual, complexitatea conţinutului şi metodologiei de elaborare a indicilor sintetici creează dificultăţi uneori insurmontabile pentru statistician, deoarece „nu există indice valabil în sine, adică indice care să nu aibă un scop" [51]. Calcularea lui trebuie să dea răspuns unei probleme centrale, adică să caracterizeze statistic esenţa fenomenelor, a căror mişcare şi schimbare o reflectă. Datele cu care se operează la elaborarea indicilor (cele efective, din numătorul raportului de comparaţie) caracterizează un aspect static al fenomenului respectiv. Dar, îndată ce le apreciem comparativ faţă de datele din perioada de referinţă, faţă de un element spaţial sau faţă de plan, care reprezintă o altă etapă sau o altă variantă a dezvoltării fenomenului studiat, noi relevăm aspectul dinamic al dezvoltării, mişcarea şi schimbarea în timp, diferenţa comparativă în spaţiu. Acesta este sensul teoretic al construirii indicilor şi în acest scop este necesară o tehnică riguroasă [85].

Sfera de cuprindere, respectiv componenţa indicilor. În forma cea mai simplă, indicii statistici se prezintă ca rapoarte între mărimile unei caracteristici sau ale unui element, înregistrate la două momente sau în două locuri diferite, ceea ce matematic se exprimă prin formulele:

$i_q = \frac{q_n}{q_0} \times 100$, pentru exprimarea variaţiei cantitative a caracteristicii;

$i_p = \frac{p_n}{p_0} \times 100$, pentru exprimarea variaţiei calitative a caracteristicii (a preţurilor).

În acest sens, dacă comparăm datele din tabelul 7.2 ale oricărui an cu unul anterior, obţinem indicii cantitativi respectivi.

Indicii simpli (elementari) se bucură de cele două calităţi intime:

— *reversibilitatea*, în sensul că variaţia poate fi calculată şi invers, adică indicele fenomenului din perioada n, raportată la baza 0, este egal cu indicele fenomenului din perioada 0 în raport cu baza t:

$$i_{n/0} \times i_{0/n} = 1; \quad i_{n/0} = \frac{1}{i_{0/n}}; \quad i_{0/n} = \frac{1}{i_{n/0}}$$

şi exemplificat cu indicatorul al treilea din tabelul 7.2:

$$i_{1971/70} = \frac{114,2}{107,3} = \frac{1}{\frac{107,3}{114,2}}; \quad i_{1970/71} = \frac{107,3}{114,2} = \frac{1}{\frac{114,2}{107,3}};$$

Scanned with OKEN Scanner

— *transferabilitatea*, prin care se înțelege că, schimbînd baza, indicele transferat este egal cu indicele inițial împărțit la raportul dintre baza transferată și baza inițială sau înmulțit cu raportul dintre baza inițială și baza transferată, adică:

$$i_{n/0} = i_{1/0} \cdot i_{2/1} \cdot i_{3/2} \dots i_{n/n}.$$

Transferabilitatea permite schimbarea bazei de comparare a indicilor. Astfel, folosind exemplul de mai sus

$$i_{1971/70} = \frac{114,2}{107,3} = 1,064 \text{ sau } 106,4\%; \quad i_{1972/71} \frac{115,5}{114,2} = 1,012 \text{ sau } 101,2\%,$$

iar

$$i_{1972/70} = \frac{114,2}{107,2} \times \frac{115,6}{114,2} = \frac{115,6}{107,3} = 1,077 \text{ sau } 107,7\%,$$

adică

$$1,064 \times 1,012 = 1,077 \text{ sau } 107,7\%.$$

Transferabilitatea nu constituie o condiție pentru orice mod de construire a indicilor de grup (sintetici), care — după cum am spus — caracterizează variația unei mărimi complexe, alcătuită din mărimi simple, cu ponderi diferite. Construirea lor implică o alegere între mai multe procedee de calcul și conduce, astfel, la cîteva forme de indici sintetici.

Indicii agregați constituie grupa indicilor sintetici cu vechimea și răspîndirea cea mai mare din istoria statisticii. Construirea formei agregate a acestor indici are la bază datele inițiale ale fiecărui element al fenomenului complex, după cum se exemplifică în tabelul 7.9. Datele arată creșterea cantității vîndute din fiecare marfă și reducerea prețului de vînzare de la o perioadă la alta. Se pune problema să se determine creșterea medie a volumului fizic al circulației totale și reducerea medie a prețurilor celor două mărfuri întrunite în grup.

Tabelul 7.9

CANTITĂȚI DE MĂRFURI VÎNDUTE ÎN DOUĂ PERIOADE ȘI PREȚURILE LOR
(date convenționale)

Produse	U.m.	Perioada de bază		Perioada curentă	
		Cantități vîndute Q_0	Preț (lei) P_0	Cantități vîndute Q_1	Preț (lei) P_1
A	m	800	10	1000	9
B	kg	500	5	600	3

La elaborarea indicelui în această formă se pune, în primul rînd, problema comensurabilității cantităților, respectiv a prețurilor. Este limpede că determinarea schimbării medii a volumului fizic al circulației mărfurilor nu este

Scanned with OKEN Scanner

posibilă prin însumarea cantităţilor celor două mărfuri vîndute din perioada curentă, apoi a celor din perioada de bază şi raportarea primei sume la a doua, în scopul obţinerii indicelui sintetic. Pot fi însă adunate valorile lor şi în acest caz preţurile vor reprezenta unitatea de măsură comună. Dacă se evaluează cantităţile de mărfuri din fiecare perioadă în preţurile respective şi se calculează indicele, aceasta va arăta schimbarea valorii circulaţiei mărfurilor, datorită schimbării concomitente a cantităţii şi a preţurilor:

$$I_v = \frac{\Sigma Q_1 P_1}{\Sigma Q_0 P_0} = \frac{(1\,000 \times 9) + (600 \times 3)}{(800 \times 10) + (500 \times 5)} = \frac{10\,800}{10\,500} = 1{,}0285 \text{ sau } 102{,}9\%.$$

Circulaţia mărfurilor în preţuri efective a crescut în perioada raportată de 1,03 ori sau cu 2,9% în comparaţie cu perioada de bază. Este evident că un asemenea indice nu răspunde la întrebarea: cu cîte procente a crescut, în medie, volumul fizic al circulaţiei mărfurilor? În acest scop este necesar să se facă abstracţie de faptul că s-au schimbat preţurile şi să se folosească pentru ambele perioade aceleaşi preţuri, fie din perioada de bază, fie din perioada curentă, care să servească la construirea indicelui volumului fizic drept element de ponderare.

Dacă folosim în acest scop preţurile din perioada de bază, indicele variaţiei cantitative, pentru întregul volum de mărfuri, va fi:

$$I_Q = \frac{\Sigma Q_1 P_1}{\Sigma Q_0 P_0} = \frac{(1\,000 \times 10) + (600 \times 5)}{(800 \times 10) + (500 \times 5)} = \frac{13\,000}{10\,500} = 1{,}238 \text{ sau } 123{,}8\%,$$

adică volumul fizic al circulaţiei a sporit în medie cu 23,8%.

Pentru calcularea indicelui variaţiei calitative (preţurilor) se pot întruni laolaltă preţurile diferitelor mărfuri vîndute, folosindu-se de obicei drept ponderi cantităţile acestora din perioada curentă. Deci, în exemplu,

$$I_P = \frac{\Sigma P_1 Q_1}{\Sigma P_0 Q_1} = \frac{(9 \times 1\,000) + (3 \times 600)}{(10 \times 1\,000) + (5 \times 600)} = \frac{10\,800}{13\,000} = 0{,}8307 \text{ sau } 83{,}1\%.$$

ceea ce înseamnă că, în medie, preţurile s-au redus cu 16,9%.

Caracteristica indicilor agregaţi este măsurarea mărimilor pentru care se calculează, prin înmulţirea lor cu coeficienţii de măsurare comună (ponderi). Condiţia care dă posibilitatea construirii formei agregate a indicelui o constituie existenţa datelor, în expresie naturală, referitoare la cantitatea producţiei fabricate, a mărfurilor vîndute etc.

Indicii de forma mediei aritmetice şi a mediei armonice. În unele cazuri datele sînt prezentate numai în expresie bănească şi atunci nu este posibilă calcularea indicelui în forma agregată. Trebuie să se recurgă la construirea indicilor sintetici, ca medii ale indicilor individuali (elementari). La întocmirea acestor indici se pune problema alegerii corecte a formei de medie şi a sistemului de ponderare. Această problemă se rezolvă pe baza regulii generale: indicele agregat constituie forma de bază a oricărui indice. Regula este respectată atunci cînd indicele calculat ca medie a indicilor individuali este egal cu indicele agregat [88].

Scanned with OKEN Scanner

Cînd dispunem de datele asupra valorii producției în perioada de bază și de indicii individuali ai volumului fizic al producției, stabilim valoarea producției în perioada curentă prin calcularea *indicelui mediu aritmetic*, care este o formă transformată a indicelui agregat. Ca exemplificare folosim datele din tabelul 7.10.

Tabelul 7.10

VARIAȚIA VOLUMULUI FIZIC AL UNEI GRUPE DE PRODUSE ÎNTRE DOUĂ PERIOADE

(date convenționale)

Produse	Valoarea producției în perioada de bază (mii lei) $Q_0 P_0$	Indicii volumului fizic al producției $i_Q = \frac{Q_1}{Q_0} \cdot 100$
A	400	110
B	200	105
C	500	108
Total	1100	108,2

În acest caz nu se cunoaște mărimea convențională Q_1P_0, care constituie numărătorul indicelui agregat. Pe baza elementelor cunoscute:

$$i_Q = \frac{Q_1}{Q_0}; \quad Q_1 = iQ_0, \text{ de unde}$$

$$Q_1P_0 = iQ_0P_0;$$

deci:

$$I_{Q_{1/0}} = \frac{\Sigma iQ_0P_0}{\Sigma Q_0P_0}$$

$$I_{Q_{1/0}} = \frac{(1{,}1 \times 400) + (1{,}05 \times 200) + (1{,}08 \times 500)}{400 + 200 + 500} = \frac{1190}{1000} = 1{,}0818 \text{ sau } 108{,}2\%.$$

Prin urmare, indicele mediu aritmetic calculat reprezintă media aritmetică a indicilor individuali, ponderați — în exemplul dat — cu valoarea producției din perioada de bază.

Egalitatea cu indicele agregat al volumului fizic al producției rezultă din relația:

$$\frac{\Sigma iQ_0P_0}{\Sigma Q_0P_0} = \frac{\Sigma \frac{Q_1}{Q_0} Q_0P_0}{\Sigma Q_0P_0} = \frac{\Sigma Q_1P_0}{\Sigma Q_0P_0}$$

și această egalitate este în funcție de un anumit sistem de ponderare.

Scanned with OKEN Scanner

Să calculăm acum indicele sintetic al prețurilor, pe baza datelor din tabelul 7.11 [88].

Tabelul 7.11

VARIAȚIA PREȚULUI UNEI GRUPE DE PRODUSE ÎNTRE DOUĂ PERIOADE

(date convenționale)

Produsul	Valoarea cantităților vîndute		Indicii prețurilor
	în perioada de bază Q_0 P_0	în perioada curentă Q_1 P_1	$i_p = \frac{P_1}{P_0} \cdot 100$
A	8 000	9 000	90,0
B	2 500	1 800	60,0
Total	10 500	10 800	83,1

Pentru calcularea indicelui prețurilor în forma agregată sînt necesare date asupra circulației efective a mărfurilor în perioada curentă P_1Q_1 și asupra circulației convenționale a mărfurilor P_0Q_1. În exemplul luat nu se cunoaște P_0Q_1, dar se pot calcula, pe baza circulației mărfurilor în perioada curentă, indicii corespunzători ai prețurilor:

$$i_p = \frac{P_1}{P_0}; P_0 = \frac{P_1}{i_p} = \frac{1}{i_p} P_1, \text{ de unde } I_{P_{1/0}} = \frac{\Sigma P_1Q_1}{\Sigma \frac{1}{I_P} P_1Q_1}$$

$$I_{P_{1/0}} = \frac{9\,000 + 1\,800}{\frac{1}{0,9} \times 9\,000 + \frac{1}{0,6} \times 1800} = 0,831 \text{ sau } 83,1\%.$$

Am folosit *indicele mediu armonic*, care reprezintă o formă transformată a indicelui agregat:

$$\frac{\Sigma P_1Q_1}{\Sigma \frac{1}{i} P_1Q_1} = \frac{\Sigma P_1Q_1}{\Sigma \frac{1}{\frac{P_1}{P_0}} P_1Q_1} = \frac{\Sigma P_1Q_1}{\Sigma \frac{P_0}{P_1} P_1Q_1} = \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1},$$

relație care confirmă că indicii prețurilor, calculați după forma armonică și agregată, coincid.

Indicele mediu armonic al prețurilor se poate calcula numai în sistemul de ponderare cu cantitățile din perioada curentă și va fi egal cu indicele agregat numai în cazul în care ponderile indicilor individuali vor fi constituite din numărătorul indicelui agregat. Din această regulă generală rezultă următoarele: indicele mediu armonic trebuie folosit numai în acele cazuri în care numărătorul indicelui agregat este o valoare cunoscută.

Indicele valorilor medii. O altă modalitate de elaborare a indicilor sintetici constă în calcularea unui raport între două mărimi, atunci cînd este vorba de o colectivitate împărțită în grupe, avînd valorile caracteristicii în raport cu

Scanned with OKEN Scanner

aceste grupe x'_o; x''_o; ..., x^n_o, în perioada de bază și x'_1, x''_1, ..., x^n_1, în perioada curentă, iar volumul grupelor f'_o, f''_o, ..., f^n_o și, respectiv, f'_1, f''_1, ...f^n_1.

Indicii caracteristici rezultă din relația

$$I_{x_{1/0}} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0}.$$

Pentru exemplificare vom folosi datele din tabelul 7.12.

Tabelul 7.12

VARIAȚIA RETRIBUȚIEI MEDII LA UN ANSAMBLU DE CATEGORII DE PERSONAL

(date convenționale)

Categorii de personal	Perioada de bază		Perioada curentă	
	Fondul de retribuire a muncii (lei) S_0	Numărul mediu scriptic T_0	Fondul de retribuire a muncii (lei) S_1	Numărul mediu scriptic T_1
A	108 000	90	132 000	100
B	67 500	45	97 200	60
C	27 000	15	75 600	40
Total	202 500	150	304 800	200

Dinamica retribuției medii, pe ansamblul categoriilor de personal, este caracterizată de indicele:

$$I_{s1/0} = \frac{\Sigma S_1}{\Sigma T_1} : \frac{\Sigma S_0}{\Sigma T_c} = \frac{304\,800}{200} : \frac{202\,500}{150} = \frac{1\,524}{1\,360} = 1{,}1288 \text{ sau } 112{,}9\%.$$

Dacă se compară indicele sintetic obținut cu indicii pentru fiecare categorie de personal, egali cu

$$i_A = \frac{132\,000}{100} : \frac{108\,000}{90} = \frac{1\,320}{1\,200} = 1{,}1 \text{ sau } 110\%,$$

$$i_B = \frac{97\,200}{62} : \frac{67\,500}{45} = \frac{1\,620}{1\,500} = 1{,}08 \text{ sau } 108\%,$$

$$i_C = \frac{75\,600}{40} : \frac{27\,000}{15} = \frac{1\,890}{1\,800} = 1{,}05 \text{ sau } 105\%,$$

observăm că la prima categorie de personal retribuția medie a crescut cu 10%, la cea de a doua cu 8%, iar la a treia categorie de personal cu 5% și, cu toate acestea, pe totalul celor trei categorii de personal, retribuția medie a crescut cu 12,9%.

Un asemenea *paradox statistic* se întîlnește de multe ori în practica calculului indicilor. Se întîmplă și cazuri inverse, în care indicele general este mai mic decît cel mai mic indice individual. Acest fapt denotă că indicele general reflectă nu numai dinamica valorilor caracteristicii, ci și influența unui factor

Scanned with OKEN Scanner

suplimentar. Este vorba, în exemplul dat, de modificarea greutății specifice a numărului de personal în perioada curentă față de perioada de bază (tabelul 7.13).

Tabelul 7.13

VARIAȚIA GREUTĂȚII SPECIFICE A CATEGORIILOR DE PERSONAL

(date convenționale)

Categorii de personal	Numărul mediu scriptic al personalului în % față de total	
	Perioada de bază	Perioada curentă
A	60	50
B	30	30
C	10	20
Total	100	100

Se constată o dublare a greutății specifice a personalului ultimei categorii la care se înregistrează retribuția medie cea mai ridicată, ceea ce explică și valoarea superioară a indicelui general în raport cu indicii individuali.

Formula de calcul scrisă sub forma (S fiind egal cu $\bar{s}T$)

$$I_{\bar{s}_{1/0}} = \frac{\Sigma \bar{s}_1 T_1}{\Sigma T_1} : \frac{\Sigma \bar{s}_0 T_0}{\Sigma T_0},$$

exprimă cu claritate semnificația indicelui, ca raport între două medii aritmetice cu ponderi diferite, mărimea lui depinzînd nu numai de schimbarea pură a retribuției medii, ci și de schimbarea structurii personalului. Avem aici un *indice cu structură variabilă.*

Paradoxul aparent se elimină dacă, în calculul indicelui, se ia, pentru ambele perioade, aceeași structură a personalului, adică:

$$I_{\bar{s}_{1/0}} = \frac{\Sigma \bar{s}_1 T_0}{\Sigma T_0} : \frac{\Sigma \bar{s}_0 T_0}{\Sigma T_0} = \frac{\Sigma \bar{s}_1 T_0}{\Sigma \bar{s}_0 T_0}.$$

În exemplul nostru avem (după tabelul 7.12):

$$I_{\bar{s}_{1/0}} = \frac{(1\,320 \times 90) + (1\,620 \times 45) + (1\,890 \times 15)}{(1\,200 \times 90) + (1\,500 \times 45) + (1\,800 \times 15)} = \frac{220\,050}{202\,500} = 1{,}0866 \text{ sau } 108{,}7\%.$$

După cum se vede, indicele obținut se situează între cel mai mic (105%) și cel mai mare indice individual (110%).

Cu ajutorul acestui indice, numit *indice cu structură fixă*, se caracterizează dinamica pură a valorilor caracteristicii, respectiv retribuția medie, neafectată de schimbarea structurii colectivității.

Gradul de influență a schimbărilor în structura colectivității asupra dinamicii retribuției medii se exprimă în *indicele de structură*, definit în relația:

$$I_{T_{1/0}} = \frac{\Sigma \bar{s}_1 T_1}{\Sigma T_1} : \frac{\Sigma \bar{s}_1 T_0}{\Sigma T_0}.$$

Scanned with OKEN Scanner

Această formulă de calcul se deduce din indicele cu structură variabilă, luînd numărul unităţilor colectivităţii pe grupe din perioadele respective, iar valoarea caracteristicii la nivelul din perioada curentă.

În exemplul folosit, ca urmare a schimbării structurii colectivităţii, retribuţia medie pentru întregul personal a sporit cu 3,9%, adică:

$$I_{T_{1/0}} = \frac{304\,800}{200} : \frac{220\,050}{150} = 1{,}0388 \text{ sau } 103{,}9\%.$$

Între ultimii trei indici există următoarea legătură:

Indicele cu structură variabilă = Indicele cu structură fixă × Indicele schimbării structurii,

respectiv

$$112{,}9 = (1{,}087 \times 1{,}039) \times 100.$$

Din relaţia existentă rezultă că oricare dintre cei trei indici poate fi calculat pe baza valorilor celorlalţi doi.

Sistemul de ponderare. Din cele arătate pînă acum se desprinde importanţa tehnicii de *ponderare*, care constituie una din problemele-cheie ale metodologiei de construire a indicilor statistici şi, după cum am menţionat şi mai înainte, a însemnat una din primele probleme ce s-au pus în istoria acestei metodologii, mai ales de cînd Arthur Young a introdus acel coeficient k, pentru a marca importanţa relativă a diferitelor mărfuri la calcularea indicelui de grup al preţurilor.

Metoda de ponderare capătă o bază ştiinţifică prin propunerea făcută în 1864 de Étienne Laspeyres, statistician german (1834—1913), de a se determina coeficientul k după cantităţile produse, vîndute sau consumate şi în mod cu totul special să se ia în considerare cantităţile din perioada de bază. Astfel a apărut formula indicelui care în literatura statistică poartă denumirea de *metoda lui Laspeyres* sau *indicele Laspeyres*:

$$I = \frac{pQ}{PQ}, \quad \text{respectiv } I = \frac{P_1Q_1}{P_0Q_0}.$$

Procedeul lui Laspeyres este utilizat şi azi pe o scară foarte largă în statistica ţărilor capitaliste. Statistica generală a Franţei, de pildă, îl foloseşte actualmente la calcularea indicelui ponderat al preţurilor de gros pentru 350 de mărfuri.

Formula lui poate fi scrisă şi astfel:

$$I = \frac{p_0Q_0 \dfrac{p_1}{p_0}}{p_0Q_0}$$

fiind o medie ponderată a unui raport de preţuri. Coeficienţii de ponderare sînt proporţionali cu valorile corespunzătoare din perioada de bază. Aceşti coeficienţi rămîn constanţi cîtă vreme nu se schimbă baza. Formula lui Laspeyres a fost extinsă şi la calcularea indicelui volumului fizic, adică

$$I = \frac{\Sigma Q_1P_0}{\Sigma Q_0P_0}$$

ponderea formînd-o, în acest caz, preţurile din perioada de bază.

Scanned with OKEN Scanner

Mai tîrziu, în 1874, Hermann Paasche, economist german (1831—1925), a propus un indice, utilizat în general şi azi, prin care se compară preţurile, pe baza cantităţilor articolelor celor mai noi intrate în consum. Formula propusă de Paasche este

$$I = \frac{\Sigma p_1 Q_1}{\Sigma p_0 Q_1}, \quad \text{scrisă şi } I = \frac{\Sigma pQ \frac{p_1}{p_0}}{\Sigma p_0 Q_1}.$$

În formula lui Paasche, ponderea se schimbă la fiecare calculare a indicelui; în felul acesta indicele devine mai reprezentativ dar schimbarea continuă a ponderii complică calculul. Şi formula lui Paasche a fost ulterior extinsă la elaborarea indicelui volumului fizic.

Cum este şi normal, indicii calculaţi după formulele lui Laspeyres şi Paasche nu dau aceleaşi rezultate decît în cazuri cu totul excepţionale. Ca atare, cercetările pentru găsirea unei metode mai bune au continuat, dar statistica ţărilor capitaliste n-a reuşit, din acest punct de vedere, decît să găsească noi formule de compromis. Astfel, Y. F. Edgeworth a propus pentru coeficientul k formula:

$$k = \frac{Q + q}{2},$$

în care Q şi q reprezintă cantităţile din perioada curentă şi perioada de bază. Indicele va avea deci formula:

$$I = \frac{\Sigma p_1 (Q_1 + Q_0)}{\Sigma p_0 (Q_1 + Q_0)},$$

ponderea reprezentînd astfel o medie a cantităţilor din cele două perioade. O altă formulă de compromis a fost propusă de statisticianul polonez *Jan Wisniewski*, decedat în 1945, în sensul unui indice mediu, elaborat de el cu ajutorul calculului integral, care marchează o mişcare liniară a preţurilor şi cantităţilor.

Formula de compromis cea mai renumită, care a suscitat discuţii încă nesfîrşite şi care este practic utilizată de statistica S.U.A. şi a altor ţări capitaliste, dealtfel puţine la număr, este aceea a indicelui numit *ideal*, al economistului american *Irving Fisher*, în lucrarea *The making of indice numbers*. Formula acestui indice reprezintă, în fond, o medie geometrică a formulelor lui Laspeyres şi Paasche, adică:

$$I = \sqrt{\frac{\Sigma Q_1 P_0}{\Sigma Q_0 P_0} \times \frac{\Sigma Q_1 P_1}{\Sigma Q_0 P_1}}, \quad \text{pentru indicele volumului fizic}$$

şi

$$I = \sqrt{\frac{\Sigma P_1 Q_0}{\Sigma P_0 Q_0} \times \frac{\Sigma P_1 Q_1}{\Sigma P_0 Q_1}}, \quad \text{pentru indicele preţurilor,}$$

În timp ce la Laspeyres şi Paasche formulele propuse au la bază media aritmetică, Fisher foloseşte media geometrică, pentru care s-au pronunţat şi alţi teoreticieni, de exemplu Stanley Jevons şi Walsh. Unii statisticieni au propus, ca un procedeu mai bun, din punct de vedere al simplităţii şi rapi-

Scanned with OKEN Scanner

ditații, calcularea indicelui pe baza medianei; de pildă Bowley și Edgeworth au arătat că totuși între media aritmetică și cea geometrică este de preferat a doua [12].

În contextul diferitelor procedee de ponderare, deci de construire a indicilor sintetici, mai amintim formula matematicianului și filozofului german Moritz Wilhelm Drobisch:

$$I = \frac{\Sigma P_1 Q_1}{\Sigma Q_1} : \frac{\Sigma P_0 Q_0}{\Sigma P_0},$$

care se regăsește în indicele de structură despre care am vorbit, dar într-o altă notație a elementelor componente, corespunzătoare aplicației făcute (dinamica retribuției medii). Formula lui Drobisch poate fi scrisă și astfel:

$$I = \frac{\dfrac{\Sigma P_1 Q_1}{\Sigma Q_1}}{\dfrac{\Sigma P_0 Q_0}{\Sigma Q_0}};$$

ea continuă a fi utilizată pentru reflectarea dinamicii deplasărilor (schimbărilor) structurale.

De asemenea, menționăm și formula statisticianului francez Lucien March:

$$I = \frac{\Sigma \dfrac{P_1}{P_0} Q_1}{\Sigma Q_1},$$

care diferă fundamental de formula lui Ir. Fisher. Ca și acesta, March a formulat unele „condiții teoretice necesare" pentru elaborarea indicilor [13], plecînd tot de la analiza întreprinsă în scopul construirii formulei indicilor prețurilor [14].

Baza de comparare. În construirea indicilor, alegerea bazei de comparare nu constituie o simplă problemă tehnică de calcul, ci are un caracter și o importanță politico-economică. Profesorul Henri Guitton arată că în măsura în care viața economică înseamnă un progres, acesta se asociază cu schimbări de structură, cu evoluții instituționale, cu schimbări în psihologie și gusturi, cu necesități noi și cu progresul tehnicii; pe măsură ce lumea îmbătrînește, fiecare bază de comparare — oricît de bună ar fi — își pierde calitatea. Schimbarea bazei constituie astfel o problemă intrinsecă de construcție a indicelui [51].

În țările capitaliste, ciclul economic înglobînd faze de avînt și faze de depresiune, problema alegerii bazei de calcul prezintă o însemnătate deose-

[12] A. MARCHAL, *Économie politique et technique statistique*, ediția a 2-a, Paris, 1945.
[13] L. MARCH, *Les modes de mesure du mouvement de prix*, Metrom, Paris.
[14] Media geometrică, după cum arăta Fisher, satisface și condiția de reversibilitate. Pe lîngă aceasta și pe cea de transferabilitate. Fisher a fixat încă următoarele „condiții teoretice necesare" pentru construirea indicilor sintetici: identitate (formula trebuie să conducă la valoarea 1, dacă situațiile din cele două perioade sînt identice); omogenitate (indicele trebuie să fie independent de unitățile de măsură ale cantităților); proporționalitate (dacă raporturile, respectiv indicii elementari au aceeași valoare, indicele obținut trebuie să fie egal cu valoarea comună).

bită pentru urmărirea fenomenelor, deoarece, dacă baza aleasă reprezintă fie o perioadă de depresiune, fie una de prosperitate, indicii obținuți prin raportare la o astfel de bază nu reflectă corespunzător evoluția fenomenului pe un timp mai îndelungat. Uneori, prin modul de alegere a perioadei de bază realitatea poate fi denaturată. Pentru a oglindi totuși evoluția reală, în statistica țărilor capitaliste se utilizează fie calcularea unei medii a mărimilor din perioada unui întreg ciclu capitalist, fie schimbarea frecventă a bazei de calcul. Dar o asemenea schimbare creează greutăți, deoarece indicii nu mai pot fi comparați; de aceea este necesar ca, un timp, să se folosească și indicii vechi și cei noi. Aceasta înseamnă o muncă dublă și în același timp discontinuitate, care împiedică comparația. Este deci necesar să se efectueze racordarea bazei, ceea ce este posibil numai dacă indicii satisfac condiția de transferabilitate. Am văzut că indicele aritmetic nu este transferabil. Pentru a evita recalcularea directă a indicilor la noua bază, Stanley Jevons și Alfred Marshall au propus construirea așa-numiților *indici în lanț continuu* sau *indici în lanț de raport*. Metoda constă în schimbarea automată a bazei la fiecare calcul nou al indicelui; astfel, indicele prețurilor se calculează în fiecare an, luîndu-se ca bază anul anterior, baza revizuindu-se la orice schimbare spontană a structurii.

În economia socialistă, efectele industrializării, ale progresului tehnic, ca și ale altor factori de ordin social, politic, psihologic etc., determină schimbări structurale, care reclamă, de asemenea, comparații în funcție de diferite perioade de bază. Drept urmare, statistica noastră de stat folosește ca bază de calcul al indicilor anii 1938, 1950, primii ani ai diverselor cincinale etc. În diverse comparații se ia ca bază sarcina de plan. Dealtfel și sarcinile de plan sînt formulate adesea sub forma indicilor.

După specificul anumitor fenomene se folosește și în statistica noastră drept bază de comparație media mai multor perioade (ani, luni, zile ș.a.). De pildă, în agricultură, din cauza influenței puternice ce o exercită factorii meteorologici și pentru ca variațiile acestora să nu influențeze valoarea indicelui, se stabilesc anumite perioade ca bază de comparație.

Diferențierea indicilor în *indici în formă de bază* și *indici în lanț* are valabilitate numai dacă este vorba de un număr mai mare decît două perioade. Indicele în formă de bază presupune raportarea continuă la una și aceeași bază, în timp ce indicele în lanț caracterizează schimbarea fenomenului în comparație cu perioada imediat anterioară, baza deplasîndu-se continuu.

Între indicii în formă de bază și cei în lanț este o legătură în sensul că produsul acestor din urmă, luați din perioadele succesive, este egal cu indicele de bază al ultimei perioade

$$i_{n/0} = \frac{x_1}{x_0} \cdot \frac{x_2}{x_1} \cdots \frac{x_n}{x_{n-1}} = \frac{x_n}{x_0},$$

iar cîtul fiecărui indice de bază este egal cu indicele în lanț al perioadei respective:

$$i_{2/n} = \frac{x_2}{x_0} : \frac{x_1}{x_0} = \frac{x_2}{x_1}$$

$$i_{n/n-1} = \frac{x_n}{x_0} : \frac{x_{n-1}}{x_0} = \frac{x_n}{x_{n-1}}.$$

Scanned with OKEN Scanner

Relația aceasta prezintă importanță practică, în primul rînd, pentru schimbarea bazei indicelui.

Aplicarea metodei sondajului la calcularea indicilor. Practica a dovedit, în frecvente cazuri, că obținerea indicilor statistici care reflectă dinamica fenomenelor în diverse sectoare ale vieții social-economice întîmpină dificultăți uneori insurmontabile, din cauza numărului și a varietății extrem de mari a cazurilor individuale care compun colectivitatea cercetată.

În scopul soluționării acestei probleme, care este strîns legată și de necesitatea economisirii cheltuielilor de muncă, timp și materiale, s-a recurs la utilizarea metodei sondajului în calculul indicilor statistici, această soluție găsindu-și o aplicare tot mai largă în statistica din țările socialiste.

Din teoria și practica aplicării metodei sondajului la calculul indicilor statistici se cunoaște că procedeul cel mai simplu, care poate fi utilizat, este *sondajul întîmplător repetat*, cu posibilități egale din întreaga colectivitate.

O altă metodă aplicabilă este *sondajul cu probabilități proporționale*, care constă în organizarea operațiilor într-un astfel de mod încît posibilitatea pentru fiecare unitate de a fi extrasă din colectivitatea generală să fie egală cu valoarea totală. Aceasta înseamnă că, înainte de a se proceda la operația de selecție, trebuie să se cunoască atît întregul nomenclator de mărfuri, cît și valoarea lor totală.

O a treia metodă, care întrunește tot mai multe sufragii și care s-a dovedit, practic, drept cea mai eficientă în diverse domenii de cercetare, de pildă în circulația mărfurilor, este așa-numita *selecție concentrată*. Aceasta constă în cuprinderea în colectivitatea de selecție a acelor elemente din colectivitatea generală, care au o pondere mai mare, respectiv care constituie elemente reprezentative în cadrul colectivității generale sau a grupelor de elemente din această colectivitate generală. De pildă, pentru grupele de produse sau mărfuri de același tip sau de aceeași natură se consideră că variațiile de preț se manifestă, în linii mari, similar cu variațiile prețurilor unora din produsele sau mărfurile care compun grupa respectivă sau cu produse asemănătoare din punctul de vedere al materialelor din care sînt confecționate și al proporției dintre prețul acestor materiale și prețul manoperei necesare pentru confecționarea lor.

Plecînd de la formula generală de calcul al indicelui de grup al prețurilor,

$$I = \frac{\Sigma P_1 Q}{\Sigma P_0 Q},$$

la calcularea indicilor, prin *metoda reprezentativității*, această formulă devine

$$I = \frac{\sum_k \sum_i p_1 k_i q_{k_i}}{\sum_k \sum_i p_{0k_i} q_{k_i}},$$

în care:

k reprezintă grupa respectivă:
i — produsele individuale din grupa k.

Formulele utilizate în acest calcul (indicele agregat, mediile aritmetică și armonică) se vor alege în funcție de gradul de apropiere dintre cantitățile, prețurile, respectiv valorile produselor (grupelor). [15]

Eficacitatea calculării indicilor prin metoda sondajului sporește dacă se utilizează procedeul stratificării colectivității generale. Acest procedeu constă în împărțirea colectivității pe straturi și calcularea indicelui în etape. Astfel, mai întîi, se calculează indicele pentru fiecare strat cu ajutorul metodei agregate. Apoi, din indicii obținuți în felul acesta, se calculează indicele mediu, folosindu-se ca pondere valoarea totală pe straturi, în colectivitățile de grup și în colectivitatea generală. Stratificarea prezintă o eficiență și mai mare în cazul existenței unei tendințe diferite a dinamicii prețurilor sau volumului fizic în diversele grupe ale colectivității cercetate, micșorîndu-se sau eliminîndu-se astfel erorile de reprezentativitate. Sondajul stratificat prezintă un avantaj atunci cînd în cercetarea statistică se urmărește și calcularea indicilor de grup.[16]

În practica aplicării metodei sondajului la calculul indicilor s-au făcut eforturi pentru soluționarea problemei determinării erorii de reprezentativitate. Ödön Éltető [17] a propus următoarea formulă pentru determinarea erorii în cazul selecției concentrate:

$$\sigma_i = \sqrt{\frac{\frac{X^2}{x^2} - \sum_{k=1}^{N} x_k^2 \Big/ \sum_{k=1}^{n} x_k^2}{\sum_{k=1}^{n} x_k^2 \Big/ x^2} \sum_{k=1}^{n} \frac{x_k^2}{X^2} (i_k - i)^2},$$

în care:

N reprezintă numărul tuturor mărfurilor;

n — numărul mărfurilor incluse în selecție.

După formula lui Laspeyres,

$x_k = q_{k_0} p_{k_0}$ este valoarea totală a mărfii k, calculată pe baza cantităților din perioada curentă și a prețurilor din perioada de bază;

$i_k = \frac{p_{k_1}}{p_{k_0}}$ — indicele individual al prețurilor mărfii k;

$i = \frac{\sum_{k=1}^{n} q_{k_0} p_{k_1}}{\sum_{k=1}^{N} q_{k_0} p_{k_0}}$ — indicele general al prețurilor, calculat pe baza datelor selecției.

[15] O tratare mai amplă a se vedea la: R. STRUCK, *Utilizarea reprezentativității în calcularea indicilor compuși*, „Statistische Praxis" nr. 7/1964.

[16] Lucrări în materie: GH. MIHOC și V. URSEANU [85]; L. TÖVISSI, *Probleme teoretice ale calculului indicilor statistici*, disertație susținută la A.S.E., București, 1965; S. DINCULESCU, GH. NISTORESCU, V.V. DUMITRESCU, *Aplicarea metodelor de selecție la elaborarea indicilor prețurilor cu ridicata*, Consfătuirea științifică cu tema „Anchetele selective", organizată de Direcția centrală de statistică, în 28—30 mai 1964, anexă la „Revista de statistică" nr. 7/1964; culegerea *Metoda selectivă în statistica curentă*, ciclul de studii regional, O.N.U., Budapesta, 16—17. IX. 1963 (trad. D.C.S., 1964); R. STRUCK, art. cit.,; I. NEGURĂ, V. URSEANU, Utilizarea *metodei selective la calcularea indicilor statistici*, culegerea *Studii de statistică*, Direcția centrală de statistică, 1970.

[17] *A reprezentativ modszerrel nyert árindex hibájának számitása* (Determinarea erorilor la calcularea indicilor prin metoda reprezentativă), „Statisztikai Szemle", februarie 1959.

Scanned with OKEN Scanner

La construirea acestei formule autorul a plecat de la următoarele condiții:

— indicele general al prețurilor (I) este considerat drept indicatorul unei colectivități cu un număr finit de unități;

— indicii individuali de prețuri ai mărfurilor cuprinse în eșantion (i_x) sînt considerați ca variabile întîmplătoare independente, avînd una și aceeași dispersie;

— ponderile indicilor individuali folosite la calcularea indicelui general sînt considerate drept mărimi constante din colectivitatea generală, iar valoarea lor nu depinde de rezultatul selecției.

La calcularea indicilor în forma agregată, mărimea distorsiunilor este influențată mai ales în legătura dintre ponderi și indicii individuali ai prețurilor. Pentru a aprecia gradul de exactitate al rezultatelor obținute prin aplicarea metodei selective, precum și, în general, pentru o analiză mai adîncită a rezultatelor invederate prin indicii obținuți, este utilă aplicarea relației construite de Bortkiewicz [18], în care indicele prețurilor este ponderat, pe rînd, cu cantitățile din perioada curentă și perioada de bază.

$$\frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} : \frac{\Sigma q_0 p_0}{\Sigma q_0 p_1} = 1 + r \cdot V_{I_p} \cdot V_{I_q},$$

în care:

I_p și I_q reprezintă indicii medii ai prețurilor și volumului fizic, ponderați cu mărimile $p_1 q_0$;

r — coeficientul liniar de corelație a indicilor prețurilor și volumului fizic, ponderați cu mărimile p_1 și q_0 și exprimat în formula

$$r = \frac{\Sigma q_0 p_1 \left(\frac{q_1}{q_0} - I_q\right)\left(\frac{p_1}{p_0} - I_p\right)}{q_0 p_1 \cdot I_p \cdot I_q}$$

σ_{I_p} și σ_{I_q} — dispersiile indicilor I_p și I_q;

V_{I_p} și V_{I_q} — coeficienții de variație ai lui I_p și I_q.

Din această relație, în cazul unei verificări a indicelui prin metoda sondajului, față de cel calculat pe baza datelor întregii colectivități, se vede că aproximarea este cu atît mai bună cu cît una din mărimile r, V_{I_p} sau V_{I_q} este mai aproape de zero, deci cu cît apropierea dintre indicii mărimilor reprezentative tinde către zero.

Cu ajutorul erorii medii de selecție σ_i putem construi intervalul în care se va găsi indicele mediu general al prețurilor, I, cu precizia probabilistică dorită. Putem aplica în acest scop inegalitatea lui Cebîșev

$$p[\,|\,i - I\,|\,] \leqslant k\sigma_i \geqslant 1 - \frac{\sigma_i^2}{k^2\sigma^2},$$

care ne indică limita inferioară a probabilității ca abaterile — în valoare absolută — ale indicelui mediu de selecție, i, de la indicele mediu în colectivitatea generală, I, să fie cuprins într-un interval dat.

După cum se știe, dispersia unei estimații este ușor de stabilit în cazul în care există o egalitate între raportul dintre numărul elementelor din colecti-

[18] *Ladislaus von Bortkiewicz* (1868—1931), statistician, matematician, economist și demograf, de origine germană, a fost profesor de economie și statistică la Petersburg și Berlin și a adus contribuții la teoria indicilor și calculul dispersiei.

vitatea de selecţie şi al celor din colectivitatea generală $\left(\frac{n}{N}\right)$ şi raportul dintre valoarea circulaţiei, de pildă, în selecţie şi valoarea circulaţiei totale $\left(\frac{v}{V}\right)$. Această condiţie nu este dată la selecţia cu probabilităţi proporţionale, întrucît $\left(\frac{v}{V}\right)$ este mult mai mare decît $\left(\frac{n}{N}\right)$. Formula propusă de Éltetö ţine seama de diferenţa dintre cele două raporturi.

7.3.3 Indicii de structură şi descompunerea influenţelor pe factori

Noţiunile teoretice dobîndite pînă acum şi procedeele de construire a indicilor sintetici, în special sistemul de ponderare, ne permit să pătrundem mai adînc în componenţa acestor indici pentru a analiza influenţa schimbărilor de structură asupra dinamicii caracteristicii.

Indicele schimbărilor (deplasărilor) structurale, numit pe scurt indice de structură, măsoară gradul repercusiunilor determinate de schimbările din structura colectivităţii asupra dinamicii caracteristicii. El este un indice factorial al indicelui cu structură variabilă, ca şi indicele cu structură fixă şi reflectă efectele schimbărilor de structură nu în sine, ci asupra evoluţiei mediei caracteristicii. El se determină totdeauna pe baza valorii acestei caracteristici (productivitatea muncii, preţul de cost, retribuţia medie ş.a.), în raport cu care interesează efectul deplasărilor structurale din sînul colectivităţii. Mai trebuie precizat că, reflectînd întreaga influenţă a schimbării structurii colectivităţii asupra dinamicii mediei caracteristicii, în această influenţă se cuprinde şi efectul evoluţiei simultane (concomitente), interdependente, a schimbării structurii şi a dezvoltării caracteristicii însăşi (interacţiunea în mişcare) [12].

După cum s-a arătat şi mai înainte, statistica dispune de variate metode de măsurare a influenţelor ce le exercită diverşi factori asupra modificării unui fenomen. Astfel sînt indicatorii corelaţiei, coeficientul de regresie ş.a. În cazul corelaţiei, privind aspectul static al legăturii dintre factori, efectuarea prealabilă a grupării permite descompunerea dispersiilor. La aplicarea metodei sondajului în calcularea indicilor am văzut cum sînt utilizaţi, în relaţia construită de *L. v. Bortkiewicz*, coeficienţii de variaţie şi corelaţie, precum şi dispersiile indicilor preţurilor şi volumului fizic.

Metoda indicilor deschide largi posibilităţi de măsurare a influenţei factorilor care, din punctul de vedere al analizei economice, prezintă o indiscutabilă importanţă. [19] Ea rezultă din faptul care apare clar în formula unui

[19] În anii '50 şi' 60 au apărut în ţara noastră şi în alte ţări socialiste numeroase studii, cursuri universitare, volume speciale sau articole de periodice, care au îmbogăţit teoria indicilor, prin aprofundarea procedeelor de descompunere a lor în factori, procedee aplicate mai ales la indicele productivităţii muncii, al preţului de cost, al retribuţiei medii, al circulaţiei mărfurilor etc. Menţionăm în special, din acest punct de vedere, pe acelea de care ne-am folosit şi aici: GH. MIHOC şi V. URSEANU [85], *Bazele teoriei generale a statisticii* [12]; V. TREBICI, *Determinarea influenţei schimbării productivităţii muncii şi a timpului chelluit asupra sporului producţiei*, „Revista de statistică" nr. 2/1956; L. TÖVISSI, *Aplicarea coeficientului de corelaţie în teoria indicilor*, „Revista de statistică" nr. 4/1961, I. NĂDEJDE şi I. NEGURĂ, *Aplicaţii ale metodei indicilor în statistica comerţului exterior*, „Revista de statistică" nr. 1/1962; L.S. KAZINEŢ, *Coeficientul de corelaţie şi problemele teoriei indicilor*; V. N. PEREGUDOV, *În legătură cu teoria metodei indicilor*, ambele studii în „Ucionîie zapiski po statistike", vol. I, Moscova, 1955; P. KÖVES, *Statisztikai indexek*, Budapesta, 1956.

indice, că un fenomen complex poate fi prezentat ca un produs de factori într-o ecuație dimensională și că schimbarea mărimii fenomenului complex este însoțită de schimbarea concomitentă a factorilor respectivi, în sensuri și cu intensități diferite. Se știe, de pildă, că valoarea unei cantități de marfă în expresie bănească este funcție de prețul unitar și de cantitate, după formula $V = f(p \cdot q)$ și, în mod asemănător, volumul producției unui produs — funcție de forța productivă a muncii (w) și timpul total de muncă cheltuit pentru producerea ei (T), ceea ce se exprimă prin formula $Q = f(w, T)$. Această relație se păstrează și sub raport dinamic.

Este evident că sporul format de la o perioadă la alta, care în cifre absolute este dat de diferența dintre valoarea produsului de la numărător și valoarea de la numitor, rezultă din schimbarea concomitentă a ambilor factori. De exemplu, la circulația unui grup de mărfuri, printr-un procedeu matematic foarte simplu, creșterea volumului circulației

$$\Sigma q_1 p_1 - \Sigma q_0 p_0$$

poate fi exprimată sub forma

$$\Sigma q_1 p_1 - \Sigma q_0 p_0 = (\Sigma q_1 p_1 - \Sigma q_1 p_0) + (\Sigma q_1 p_0 - \Sigma q_0 p_0),$$

expresie în care primul termen este realizat pe seama reducerii prețurilor, iar al doilea termen reprezintă sporul realizat pe seama creșterii cantității vîndute.

Desigur că putem avea cazuri în care sporul se formează sub influența unui număr mai mare de factori, de pildă, cheltuielile materiale sînt determinate ca produs al prețurilor unitare de achiziție, al normelor de consum și al volumului producției fabricate. Pentru a se stabili cu cît contribuie fiecare factor la formarea sporului absolut, în cazul indicilor trebuie să aplicăm ponderarea, ca proces de abstractizare. În practica statisticii din țara noastră, indicele unui element intensiv (al caracteristicii x) se ponderează cu elementul extensiv al nivelului perioadei curente (f_i), iar indicele elementului extensiv (al colectivității) se ponderează cu mărimea elementului intensiv din perioada de bază. Exemplul tipic îl constituie indicele prețurilor, care este ponderat cu cantitățile din perioada curentă și indicele volumului fizic al producției, ponderat cu prețurile din perioada de bază. Prețul este caracteristica sau elementul intensiv, iar cantitatea reprezintă factorul extensiv, x și f, conform notației de mai sus. Atît la acești indici, cît și la indicele productivității muncii în exprimare directă, sistemul de ponderare arătat are justificare economică. Dar, cum la construirea sistemului de indici cu structură variabilă, cu structură fixă și de structură au fost utilizate ca ponderi mărimile factorului extensiv din perioada de bază, nu din perioada curentă, însă *numai în scopul analizei*, tot așa la descompunerea sporurilor absolute pe factori, ponderarea pentru toți factorii se face cu mărimea factorului intensiv și extensiv din perioada de bază, obținîndu-se totodată și o parte din spor care nu se descompune pe factori, el fiind o expresie a influenței comune ce nu poate fi separată [12, 88].

Simbolizînd cu A fenomenul complex, descompunerea se va face astfel:

$$A_1 - A_0 = (x_1 - x_0)f_0 + (f_1 - f_0)x_0 + (x_1 - x_0)\ (f_1 - f_0).$$

Scanned with OKEN Scanner

Deci

$$\Delta_{A(xf)} = \Delta_{A(x)} + \Delta_{A(f)} + \Delta_{A(xf)}.$$

Un alt procedeu de descompunere este cel bazat pe teorema creşterilor finite a lui Lagrange, pentru care, dacă este vorba de doi factori, se foloseşte formula $V = f(x \cdot y)$, aceasta însă modificîndu-se astfel:

$$V(x) = v_0\alpha\left(1 + \frac{\beta}{2}\right); \quad V(y) = v_0\beta\left(1 + \frac{\alpha}{2}\right).$$

Pentru ilustrarea procedeelor arătate, folosim datele din tabelul 7.14 [20].

Tabelul 7.14

DINAMICA PRODUCTIVITĂŢII MUNCII ÎNTR-O UNITATE ECONOMICĂ

(date convenţionale)

Perioada	Volumul producţiei (Q)	Productivitatea muncii (w)	Numărul de muncitori utilizaţi (T)
De bază	2 000	20	100
Curentă	2 640	24	110

a) După metoda indicilor în lanţ:

$$\frac{Q_1}{Q_0} = \frac{W_1T_1}{W_0T_1} \cdot \frac{W_0T_1}{W_0T_0} \quad \text{şi}$$

$$Q_1 - Q_0 = (W_1 - W_0)T_1 + (T_1 - T_0)W_0 =$$
$$= (24-20)110 + (110-100)20 = 640$$

şi cu alt sistem de ponderare

$$Q_1 - Q_0 = (24-20)100 - (110-100)\,24 = 640.$$

b) După metoda determinării separate a influenţei fiecărui factor (a restului nedescompus);

$$Q_1 - Q_0 = (W_1 - W_0)T_0 + (T_1 - T_0)W_0 + (W_1 - W_0)\ (T_1 - T_0) =$$
$$= (24-20)100 + (110-100)20 + (24-20)\ (110-100) =$$
$$= 400 + 200 + 40 = 640.$$

c) Pe baza teoremei creşterilor finite (Lagrange)

$$\Delta_{Q(W)} = Q_0\alpha\left(1 + \frac{\beta}{2}\right) \quad \text{şi} \quad \Delta_{Q(T)} = Q_0\beta\left(1 + \frac{\alpha}{2}\right)$$

[20] Exemplificarea şi dezvoltarea metodologiei în [12].

Scanned with OKEN Scanner

întrucît $\alpha = \frac{W_1}{W_0} - 1$ şi $\beta = \frac{T_1}{T_0} - 1$,

$$\Delta_{Q(W)} = 2\,000 \cdot 0{,}2\left(1 + \frac{0{,}1}{2}\right) = 420$$

$$\Delta_{Q(T)} = 2\,000 \cdot 0{,}1\left(1 + \frac{0{,}2}{2}\right) = 220.$$

Se poate observa că ultimul procedeu se bazează pe principiul repartizării proporţionale a sporului pe factori.

Tabelul 7.15 sintetizează repartizarea sporului pe factori, în funcţie de procedeele arătate [12].

Tabelul 7.15

REPARTIZAREA SPORULUI ABSOLUT AL PRODUCŢIEI DUPĂ PROCEDEELE FOLOSITE

(pe baza datelor din tabelul 7.14)

Procedee de descompunere	Sporul total Q	Sporul repartizat pe seama		
		Productivităţii muncii $Q(w)$	Numărului de muncitori $Q(T)$	Acţiunii concomitente $Q(wT)$
1. Indici în lanţ				
Varianta A	640	440	200	—
Varianta B	640	400	240	—
2. Restul nedescompus	640	400	200	40
3. Procedeul creşterilor finite	640	420	220	—

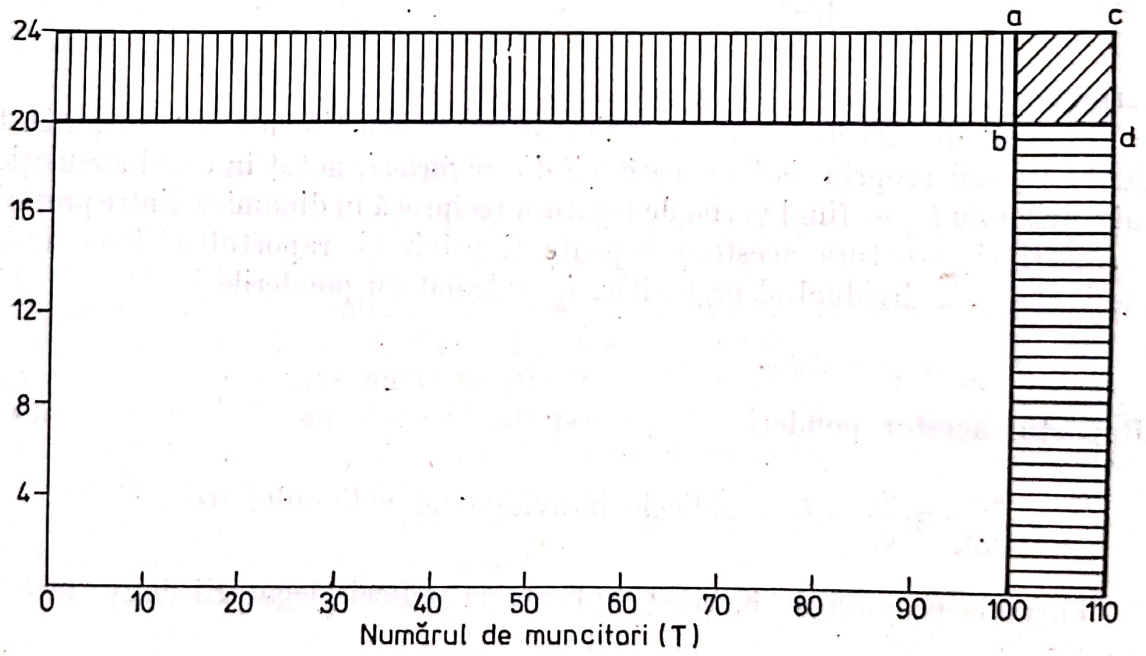


Fig. 7.9 Descompunerea sporului absolut al producţiei pe seama creşterii productivităţii muncii şi a numărului de muncitori [12].

O reprezentare grafică (fig. 7.9) scoate mai bine în evidenţă „restul nedescompus al sporului", care apare în patrulaterul *a b c d*. După primul procedeu, varianta A, acest rest este atribuit productivităţii muncii, iar în vari-

Scanned with OKEN Scanner

anta B, el revine numărului de muncitori; după procedeul al doilea, restul nedescompus rămîne neschimbat, constituind expresia influenței comune (concomitente) a celor doi factori; după cea de a treia metodă, el se împarte proporțional între productivitatea muncii și numărul de muncitori.

În exemplificarea de mai sus, sporurile determinate de cei doi factori sînt pozitive; în alte cazuri ele pot fi pozitive și negative, adică influența lor se produce în sensul creșterii din partea unuia și în sensul descreșterii din partea celuilalt. În toate cazurile metodologia descompunerii rămîne aceeași.

L.S. Kazineț [21] a propus o metodă de descompunere, plecînd de la două mărimi medii calculate cu ponderi diferite, care apar în compunerea indicelui de structură. El stabilește că raportul unor astfel de mărimi medii $\bar{x}_f$ și $\bar{x}_p$ se poate scrie:

$$\frac{\bar{x}_f}{\bar{x}_p} = \frac{\left(\bar{x} \cdot \frac{\bar{f}}{p}\right)}{\bar{x}\left(\frac{\bar{f}}{p}\right)},$$

adică media produsului dintre caracteristica x și raportul ponderilor împărțită prin media caracteristicii x înmulțită cu media raportului ponderilor.

De asemenea, diferența acelorași medii se scrie:

$$\bar{x}_f - \bar{x}_p = \frac{\overline{\left(\bar{x} \cdot \frac{\bar{f}}{p}\right)} - \bar{x} \cdot \left(\frac{\bar{f}}{p}\right)}{\left(\frac{\bar{f}}{p}\right)},$$

formulă care se deduce foarte simplu din cea precedentă.

Indicele de structură sau al schimbărilor de asortiment, care ar putea fi denumit, mai propriu, *indicele legăturii dintre factori*, notat în cazul circulației mărfurilor cu f_{pq} — fiind vorba de legătura reciprocă în dinamică dintre prețuri, cantități și structura acestora — poate fi privit ca raportul a două medii ale indicelui individual al prețurilor, i_p, calculat cu ponderile

$$f = q_1 p_0 \text{ și } p = q_0 p_0.$$

Raportul acestor ponderi

$$\frac{f}{p} = \frac{q_1 p_0}{q_0 p_0} = \frac{q_1}{q_0} = i_q = \text{indicele individual al volumului fizic.}$$

Formula propusă de Kazineț, aplicată la indicele legăturii dintre factori, ne dă

$$I_{pq} = \frac{\overline{(i_p \cdot i_q)}}{\bar{i}_p \cdot \bar{i}_q},$$

[21] L.S. KAZINEȚ, *Indicele schimbărilor structurale și coeficientul schimbării structurii*, „Naucinîie dokladî Vîșșei șkolî: Ekonomiceskie Nauki" nr. 4/1960.

expresia care se mai poate transforma, folosind definiția coeficientului de corelație a două mărimi variabile, în speță i_p și i_q, adică

$$r_{i_p i_q} = \frac{\overline{(i_p i_q)} - \bar{i}_p \cdot \bar{i}_q}{\sigma_{i_p} \sigma_{i_q}},$$

unde σ_{i_p} σ_{i_q} reprezintă abaterile medii pătratice ale celor doi indici; se obține astfel expresia transformată a indicelui legăturii dintre factori:

$$I = \frac{[\overline{(i_p i_q)} - \bar{i}_p \bar{i}_q] + \bar{i}_p \bar{i}_q}{\bar{i}_p \bar{i}_q} = 1 + \frac{\overline{(i_p i_q)} - \bar{i}_p \cdot \bar{i}_q}{\bar{i}_p \cdot \bar{i}_q} = 1 + \frac{r_{i_p i_q} \sigma_{i_p} \cdot \sigma_{i_q}}{\bar{i}_p \cdot \bar{i}_q} =$$

$$= 1 + r_{i_p i_q} V_{i_p} V_{i_q},$$

în care [22]

$$r_{i_p i_q} = \frac{\Sigma (i_p - I_{p(q_0)})(i_q - I_{q(p_0)}) q_0 p_0}{\sigma_{i_p} \sigma_{i_q} \Sigma q_0 p_0}.$$

$$\sigma_{i_p} = \sqrt{\frac{\Sigma (i_p - I_{p(q_0)})^2 q_0 p_0}{\Sigma q_0 p_0}}$$

$$\sigma_{i_q} = \sqrt{\frac{\Sigma (i_p - I_{q(p_0)})^2 q_0 p_0}{\Sigma q_0 p_0}}$$

$$V_{i_p} = \frac{\sigma_{ip}}{I_{p(q_0)}}; \qquad V_{i_q} = \frac{\sigma_{iq}}{I_{q(p_0)}}.$$

Din cele de mai sus rezultă că indicele legăturii dintre factori depinde de:

a) coeficientul de corelație dintre indicii individuali ai prețurilor și volumului fizic ($r_{i_p i_q}$);

b) coeficientul de variație a indicelui individual al prețurilor (V_{i_p});

c) coeficientul de variație a indicelui individual al volumului fizic (V_{i_q}).

Descompunerea efectuată pe baza formulelor de mai sus conduce la următoarele concluzii:

1) Indicele legăturii dintre factori este egal cu 1, dacă se anulează unul din factorii $r_{i_p i_q}$, V_{i_p}, V_{i_q}. Primul este egal cu 0, dacă și numai dacă nu există corelație liniară între indicii individuali ai prețurilor și indicii individuali ai volumului fizic; $V_{i_p} = 0$, dacă abaterea medie pătratică a indicilor individuali ai prețurilor este nulă; se știe însă că $\sigma_{i_p} = 0$, dacă toate valorile i_p sînt egale între ele, adică dacă toate prețurile perioadei curente sînt proporționale cu prețurile corespunzătoare din perioada de bază; $V_{i_q} = 0$, dacă abaterea medie pătratică a indicilor individuali ai cantităților este nulă, dar se știe că $V_{i_q} = 0$, dacă toate valorile i_q sînt egale între ele, adică dacă toate cantitățile vîndute în cele două perioade sînt proporționale.

2) Indicele legăturii dintre factori este supraunitar, dacă $r_{i_p i_q} > 0$ și $V_{i_p} \neq 0$ (coeficienții de variație nu pot lua valori negative). Condiția $r_{i_p i_q} > 0$ înseamnă că există o corelație directă între indicii individuali ai prețurilor și volumelor fizice, adică atunci cînd prețurile cresc, crește și

[22] Se utilizează simbolurile $I_{p(q_0)}$, pentru indicele general (de grup) al prețurilor cu ponderea perioadei de bază și $I_{p(q_1)}$, pentru indicele general (de grup) al prețurilor cu ponderea perioadei curente; similar, $I_{q(p_0)}$ și $I_{q(p_1)}$, pentru indicii respectivi ai volumului fizic.

Scanned with OKEN Scanner

volumul mărfurilor vîndute. Condiţiile $V_{i_p} \neq 0$ şi $V_{i_q} \neq 0$ implică inexistenţa proporţionalităţii dintre preţurile şi cantităţile perioadei de bază şi ale perioadei curente, condiţii care practic sînt realizate totdeauna.

3) Indicele legăturii dintre factori este subunitar, dacă $r_{i_p i_q} < 0$ şi $V_{i_p} \neq 0$, $V_{i_q} \neq 0$. Condiţia $r_{i_p i_q} < 0$ înseamnă că există o corelaţie inversă între indicii individuali ai preţurilor şi indicii individuali ai volumului fizic, adică atunci cînd preţurile descresc, volumul mărfurilor desfăcute creşte. Mărimea factorilor V_{i_p} şi V_{i_q} va indica în ce măsură este influenţat indicele legăturii dintre factori de fiecare din aceştia [23].

Cele de mai sus nu epuizează metoda descompunerii sporului pe factori. Numeroase propuneri au fost făcute în literatura de specialitate. Rezultate satisfăcătoare ne dă analiza regresiei, analiza factorială. Un exemplu în această privinţă ne dă funcţia de producţie de tip Cobb-Douglas:

$$Y = f(A,\ K^{\alpha}\ L^{\beta}),$$

în care:

K este stocul de fonduri fixe productive;
L — stocul de forţă de muncă;
A — parametru ce poate simboliza dimensiunea producţiei;
Y — produsul (output), ca variabilă rezultativă (produsul final sau venitul naţional);
α — coeficientul de elasticitate al variaţiei produsului în funcţie de variaţia lui K;
β — coeficientul de elasticitate al variaţiei producţiei în funcţie de variaţia lui L.

Adică:

$$\alpha = \frac{\Delta Q}{Q} = \frac{\Delta K}{K} \quad \text{şi} \quad \beta = \frac{\Delta Q}{Q} = \frac{\Delta L}{L}.$$

Cum s-a mai precizat, prioritate în toate cazurile are analiza economică care pune în evidenţă raportul real dintre factori şi rezultatul obţinut.

În cazul de mai sus, problema se reduce la a măsura contribuţia celor doi factori, precum şi aşa-numitul „efect rezidual", care poate fi atribuit progresului tehnic, influenţei perfecţionării calificării forţei de muncă etc.

7.3.4 Indicii teritoriali

Între problemele teoretice şi metodologice care s-au pus statisticii în perioada postbelică, construirea *indicilor teritoriali* s-a situat pe plan principal: într-un cadru mai general ca o problemă a statisticii internaţionale, iar din punctul de vedere al statisticii naţionale ca o cerinţă determinată de politica dezvoltării proporţionale a tuturor regiunilor unei ţări.

Indicii teritoriali formează un capitol prioritar al statisticii teritoriale, care s-a constituit ca o statistică particulară, avînd drept obiect: a) caracterizarea şi analiza nivelului repartizării teritoriale a forţelor de producţie, re-

[23] O aplicaţie a acestei metode de descompunere se găseşte la ILEANA NĂDEJDE, I. NEGURĂ, *op. cit.*

Scanned with OKEN Scanner

flectat în indicatori sintetici şi în indicatori analitici; b) aprecierea eficienţei economice a deplasărilor structurale în repartizarea teritorială a forţelor de producţie pe fiecare regiune şi comparativ între regiuni; c) analiza îndeplinirii planului de stat al economiei naţionale în profil teritorial şi aprecierea eficienţei economice a amplasării diferitelor obiective de investiţii pe regiuni; determinarea tendinţelor de dezvoltare, aprecierea statistic-economică a posibilităţilor de apropiere a nivelurilor de dezvoltare a regiunilor sub raport economic şi social-cultural.[24]

Prin asemenea obiective statistica teritorială se confundă, pe planul relaţiilor internaţionale, cu *statistica internaţională*. Sfera de investigaţie a acesteia se lărgeşte însă cu cercetarea complexă a fenomenelor economico-sociale pe spaţiul mondial, considerat ca un ansamblu teritorial, în care au loc fenomene specifice cum sînt: diviziunea internaţională a muncii, comerţul mondial, sistemul valutar şi de credit internaţional, fluxul investiţiilor internaţionale ş.a., cîmp vast în care se conturează tendinţe noi în domeniul relaţiilor economice şi culturale, cooperării tehnico-ştiinţifice şi unde apar legături şi interdependenţe între fenomene la scară mondială. Metoda indicilor teritoriali este solicitată în toată plenitudinea şi, în acest domeniu, indicii statistici dobîndesc adesea şi aspectul dinamic, comparaţia teritorială putîndu-se face şi cronologic, de pildă: indicii raportului de schimb, indicii navlului; indicele producţiei industriale mondiale, al venitului naţional, al consumului ş.a.

În acest cadru, al comparaţiilor pe plan teritorial, naţional şi internaţional, cu ajutorul metodei indicilor teritoriali se compară datele unei regiuni cu cele ale altei regiuni sau chiar datele mai multor regiuni între ele, noţiunea de regiune avînd aici un sens foarte larg: ţară, regiune administrativă, economică, politică, valutară, istorică; comparaţia se poate face şi între întreprinderi, ramuri, organizaţii, complexe economice etc. Cu alte cuvinte, prin construirea indicilor teritoriali se pune problema de a compara, pe plan spaţial, static, indicatori centralizatori cu caracter de medii.

Construirea indicilor teritoriali nu ridică probleme din punctul de vedere al formei lor, care, în principal, este aceea a indicilor agregaţi. Dar, comparaţia pe plan spaţial (şi static) implică anumite dificultăţi legate de faptul că ea se efectuează în condiţiile deosebirilor în ce priveşte structura colectivităţilor comparate, adeseori deosebit de pronunţate. Cercetările în această materie s-au pronunţat pentru rezolvarea, în prealabil, a două probleme metodologice, şi anume: 1) alegerea bazei de comparaţie şi a sistemului de ponderare; 2) asigurarea corelaţiei de sistem între indicii factoriali şi indicele fenomenului complex.[25]

Stabilirea bazei de comparare constituie aspectul principal al diferenţierii metodologiei de elaborare a indicilor teritoriali faţă de indicii cronologici şi de indicii de plan, întrucît este vorba de mai multe colectivităţi privind acelaşi proces economico-social sau mai multe grupe de aceleaşi colectivităţi

[24] SIMINA FALCAN-POPESCU, *Probleme teoretice şi metodologice ale statisticii teritoriale*, „Revista de statistică" nr. 9/1964 şi *Este statistica teritorială o disciplină relativ autonomă a statisticii?*, „Revista de statistică" nr. 1/1967.

[25] Citǎm aici din literatura noastră mai ales volumul [85], precum şi studiul elaborat de ELENA BIJI şi LUCIA PINTILIE, *Unele aspecte ale metodologiei de alcătuire a indicilor teritoriali*, „Revista de statistică", nr. 9/1962. În ce priveşte indicii teritoriali, în statistica internaţională, ne referim la [28], de asemenea la J. ZELINKA, *Comparabilitatea producţiei industriale*, traducere publicată în „Probleme de statistică social-economică" nr. 8/1958.

coexistente în același timp și pe teritorii diferite și a căror structură diferă. În aceste condiții elaborarea indicilor teritoriali trebuie să aibă în vedere numeroase caracteristici, care uneori pot varia extrem de mult. Deoarece nu totdeauna este ușor să se determine, după criterii obiective, regiunea (unitatea luată ca bază de comparare) este necesar să se facă — în vederea folosirii mediilor statistice — grupări ale indicatorilor aceluiași fenomen (proces), după caracteristici esențiale; mărimile medii calculate servesc numai pentru desemnarea grupei-bază de comparare.

În teoria statisticii, considerînd conținutul economic specific, sensul logic și funcțiile analitice ce le îndeplinesc indicii teritoriali, se face distincție între indicii teritoriali ai indicatorilor cantitativi (volumul fizic al producției, volumul fondurilor fixe și al fondurilor circulante, recolta globală ș.a.) și indicii teritoriali ai indicatorilor calitativi (productivitatea muncii, prețul de cost, gradul de utilizare a fondurilor fixe și circulante etc.).

În scopul stabilirii bazei de comparație și a sistemului de ponderi, cei mai mulți statisticieni s-au pronunțat pentru aplicarea *metodei de standardizare.*

Această metodă, elaborată cu decenii în urmă de demografie [26] pentru analiza comparativă a mortalității (între două țări, două regiuni sau două perioade), are două forme: *metoda populației standard* (metoda standardizării directe) și *metoda ratelor de mortalitate standard* (metoda standardizării indirecte).

Dacă avem două populații $P(A)$ și $P(B)$, cu mortalitățile specifice $m_{i(A)}$ și $m_{i(B)}$ și o a treia populație $P(R)$, cu mortalitățile $m_{i(R)}$, pe care o vom folosi pentru standardizare, cele două metode se înfățișează astfel:

a) metoda populației standard $P(R)$:

$$I_{m_B/m_A} = \frac{\sum_{i=0}^{\omega} m_{i(B)} P^*_{i(R)}}{\sum_{i=0}^{\omega} m_{i(A)} P^*_{i(R)}};$$

b) metoda ratelor standard de mortalitate

$$I_{m_B/m_A} = \frac{\sum_{i=0}^{\omega} m_{i(R)} P^*_{i(B)}}{\sum_{i=0}^{\omega} m_{i(R)} P^*_{i(A)}}.$$

Pentru a evita frecvențele absolute, s-au folosit cele relative

$$P^*_i = \frac{P_i}{\Sigma P_i}, \quad \text{de unde } \Sigma P^*_i = 1.$$

În primul caz, elementul de standardizare, deci elementul „fix" îl constituie structura populației R, în cel de al doilea caz — ratele de mortalitate ale populației R. Comparația devine astfel posibilă.

Prin generalizare, avînd de comparat două regiuni și luînd ca standard valorile respective pentru întreaga țară, formulele devin:

$$\bar{x}_A = \frac{\Sigma x_A f_A}{\Sigma f_A} \quad \text{sau} \quad \bar{x}_A = \Sigma x_A f^*_A,$$

[26] A se vedea [134], p. 174—176.

în care

$$f_A^* = \frac{f_A}{\Sigma f_A}$$

$$\bar{x}_B = \frac{\Sigma x_B f_B}{\Sigma f_B}, \quad \text{sau} \quad \bar{x}_B = \Sigma x_B f_B^*,$$

în care

$$f_B^* = \frac{f_B}{\Sigma f_B}.$$

Indicele $\bar{x}_{A/\bar{x}_B} = \frac{\Sigma x_A f_A^*}{\Sigma x_B f_B^*}$ este un indice a două valori medii cu ponderi diferite, deci un indice cu structură variabilă; standardizarea directă va fi

$$\bar{x}_{A/\bar{x}_B} = \frac{\Sigma x_A f_R^*}{\Sigma x_B f_R^*} \quad \text{(indice cu structură fixă),}$$

iar standardizarea indirectă:

$$\bar{x}_{A/\bar{x}_B} = \frac{\Sigma x_R f_A^*}{\Sigma x_R f_B^*} \quad \text{(indice al schimbărilor structurale).}$$

Formulele sînt prezentate în formă simplificată pentru a pune în evidență sensul celor două procedee de standardizare.

Mărimile din cele două regiuni ce se compară pot fi notate cu x_0 și x_1, respectiv l_0 și l_1; cei trei indici vor fi: I_g — indicele mărimilor medii, I_d — indicele cu structură fixă și I_{ind} — indicele schimbărilor structurale.

După cum se vede, cele două metode diferă prin faptul că la metoda directă se operează cu forma indicelui agregat (pondere egală la numărător și numitor), în timp ce la metoda indirectă este vorba de un raport între doi indici reflectînd structuri diferite ale colectivităților [27]. Din punctul de vedere al conținutului economic, I_d are o structură fixă și marchează variația indicatorului calitativ, în condițiile structurii colectivității standard. I_{ind} măsoară variația medie a schimbării structurii fiecărei grupe față de structura colectivității standard, ceea ce se exprimă prin formulele:

$$\frac{\Sigma x_1 l_1}{\Sigma x_1 l} : \frac{\Sigma x_0 l_0}{\Sigma x_0 l} = \frac{\Sigma \frac{l_1}{l_0} x_1 l}{\Sigma x_1 l} : \frac{\Sigma \frac{l_0}{l} x_0 l}{\Sigma x_0 l} \quad \text{(indicator cantitativ),}$$

respectiv

$$\frac{\Sigma x_1 l_1}{\Sigma x_1 l} : \frac{\Sigma x_0 l_0}{\Sigma x_0 l} = \frac{\Sigma \frac{l_1}{x} x l_1}{\Sigma x l_1} : \frac{\Sigma \frac{x_0}{x} x l_0}{\Sigma x l_0} \quad \text{(indicator calitativ).}$$

Se poate constata că între I_g, I_d și I_{ind} există aceeași relație pe care am întîlnit-o mai înainte, între indicele cu structură fixă și indicele de structură. Mărimea variației caracteristicii între două colectivități din spații diferite va fi determinată de influența deosebirii dintre nivelul caracteristicii (factorul calitativ) al celor două colectivități comparate și de volumul și structura lor. Folosind relația dintre indici (indicele general = indicele influenței

[27] L. S. KAZINEȚ, *O metodah postroenitia teritorialnih indeksov*, „Vestnik statistiki" nr. 3/1959; extinzînd aceste metode la studiul variațiilor sub aspect teritorial, autorul le utilizează la determinarea factorului cantitativ și a factorului calitativ.

factorului calitativ × indicele factorului cantitativ) și utilizînd, în locul ponderării cu structura colectivității standard, ponderea indicatorului care reprezintă distribuirea medie a unităților în grupe ale colectivității generale (structura medie pe regiune) [28], la determinarea, de pildă, a variației volumului fizic al producției între două colectivități, indicele teritorial (spațial) va fi dat de formula:

$$I^q_{1/0} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_0} = \frac{\Sigma w_1 T_m}{\Sigma w_0 T_m} \cdot \left(\frac{\Sigma w_1 T_1}{\Sigma w_1 T_m} : \frac{\Sigma w_0 T_0}{\Sigma w_0 T_m} \right),$$

în care:

w_0 și w_1 reprezintă productivitatea muncii a colectivităților comparate;
T_0 și T_1 — numărul muncitorilor din cele două colectivități;
T_m — structura medie regională a numărului de muncitori în colectivitatea generală.

În această formulă, structura standard a fost înlocuită cu structura medie regională din colectivitatea generală.

Inversînd baza de comparare, formula se va scrie:

$$I^q_{0/1} = \frac{\Sigma w_0 T_0}{\Sigma w_1 T_1} = \left(\frac{\Sigma w_0 T_m}{\Sigma w_1 T_m} \right) \cdot \left(\frac{\Sigma w_0 T_0}{\Sigma w_0 T_m} : \frac{\Sigma w_1 T_1}{\Sigma w_1 T_m} \right).$$

Indicii calculați cu ponderi diferite sînt reversibili. Produsul lor este egal cu 1. Ajungem astfel la tipul „indicelui ideal" al lui Irving Fischer, iar după denumirea dată de V. Peregudov — la „indicele legăturii reciproce". Acesta din urmă atrage atenția că, dacă în cazul analizei în timp cerința reversibilității este pur formală, atunci cînd regiunile comparate sînt echivalente, „reversibilitatea în spațiu" este necesară: în caz contrar, nu se va putea depăși contradicția ce apare adeseori în practică; de exemplu, diferența de structură a producției în regiunile comparate conduce la rezultate contradictorii. Cu alte cuvinte, pentru construirea indicelui respectiv se cere ca fiecare regiune să poată fi luată drept bază.[29] Pentru exemplificare utilizăm datele convenționale din tabelul 7.16.

Tabelul 7.16

ELEMENTELE DE COMPARAȚIE ALE COSTULUI UNOR PRODUSE PENTRU DOUĂ COLECTIVITĂȚI REGIONALE

Produse	Regiunea 1				Regiunea 2		Cheltuieli de producție	
	c_1	q_1	$c_1 q_1$	c_2	q_2	$c_2 q_2$	$c_1 q_1$	$c_2 q_1$
A	5	440	2 200	2,5	600	1 500	3 000	1 100
B	10	160	1 600	8	175	1 400	1 750	1 280
C	20	61	1 220	20	60	1 200	1 200	1 220
D	40	40	1 600	50	36	1 800	1 440	2 000
E	50	44	2 200	100	25	2 500	1 250	4 400
Total	—	—	8 820	—	—	8 400	8 640	10 000

[28] E. BIJI, L. PINTILIE, *op. cit.* Se poate lua însă structura pe întreaga țară, nu cea medie pe ambele colectivități.

[29] V. PEREGUDOV, *K voprosu o sravnenii po raionam kacestvennih pokazatelei*, „Vestnik Statistiki", nr. 5/1959. Prezentare în detaliu în [85], p. 439—444.

Scanned with OKEN Scanner

Calculînd indicii agregaţi respectivi, prin utilizarea ca bază a fiecăre regiuni, obţinem

$$I_{2/1} = \frac{\Sigma c_2 q_2}{\Sigma c_1 q_2} = \frac{8\,400}{8\,640} = 0{,}972 \quad \text{şi} \quad I_{1/2} = \frac{\Sigma c_1 q_1}{\Sigma c_2 q_1} = \frac{8\,820}{10\,000} = 0{,}882.$$

Comparaţia acestor doi indici cu unitatea conduce la concluzia absurdă că în a doua regiune situaţia este mai bună (preţul de cost mai redus cu 2,8%) şi totodată că prima regiune are o situaţie mai favorabilă (preţul de cost mai scăzut cu 11,8%). Dar, aplicînd media geometrică, rezultatul este:

$$I_{geom} = \sqrt{0{,}972 \cdot 0{,}882} = 0{,}926,$$

adică preţul de cost este mai scăzut în prima regiune în comparaţie cu regiunea a doua.

Întrucît din punct de vedere al conţinutului economic indicele legăturii reciproce $I(c,q)$ reprezintă specializarea relativă a regiunilor într-un sens strict al cuvîntului (în raport cu o anumită caracteristică), influenţa acestei specializări poate fi eliminată împărţind indicii $I_{2/1}$ şi $I_{1/2}$ prin indicele specializării relative, adică

$$I'_{2/1} = I_{2/1} : I_{geom} = 0{,}972 : 0{,}926 = 1{,}050$$

$$I'_{1/2} = I_{1/2} : I_{geom} = 0{,}882 : 0{,}926 = 0{,}952,$$

produsul lor fiind egal cu 1.

7.4 CALCULUL STATISTIC AL TENDINŢEI DE DEZVOLTARE ŞI MĂSURAREA OSCILAŢIILOR SEZONIERE

7.4.1 Procedee de ajustare a seriilor cronologice

Cum am arătat la începutul acestui capitol, ca efect al influenţei factorilor care determină modul de dezvoltare (evoluţie) a fenomenelor şi proceselor oglindite în seriile cronologice, mişcările şi schimbările apar sub formă de tendinţă de durată, de variaţii periodice (sezoniere), oscilaţii incidentale, iar uneori, mai ales în lumea capitalistă, şi ca variaţii ciclice. Pentru ilustrarea acestor forme de mişcare, vom exemplifica cu datele evoluţiei unui fenomen (tabelul 7.17), în evoluţia căruia au putut fi identificate influenţele unei multitudini de factori; cunoaşterea lor implică o analiză calitativă prealabilă, pe baza legităţilor ce guvernează acţiunea lor.

Scanned with OKEN Scanner

Tabelul 7.17

VARIAȚIILE PREȚULUI MEDIU ANUAL ȘI LUNAR AL GRÎULUI ÎN COMERȚUL INTERNAȚIONAL ÎN ANII 1934—1956, RESPECTIV 1955—1956

Anii	Prețuri medii anuale	Anii	Prețuri medii anuale	Lunile	Prețuri medii lunare în anii	
					1955	1956
1934	30	1946	90	Ianuarie	65	63
1935	31	1947	105	Februarie	65	64
1936	45	1948	83	Martie	66	65
1937	48	1949	74	Aprilie	65	65
1938	23	1950	74	Mai	66	65
1939	28	1951	85	Iunie	66	66
1940	25	1952	82	Iulie	66	65
1941	26	1953	70	August	66	65
1942	32	1954	65	Septembrie	65	64
1943	47	1955	65	Octombrie	63	65
1944	49	1956	64	Noiembrie	64	65
1945	52			Decembrie	63	65

Reprezentarea grafică a datelor din tabel, așa cum ne-o arată figura 7.10, ușurează urmărirea tendinței de lungă durată a mișcării, precum și a diferitelor variații.

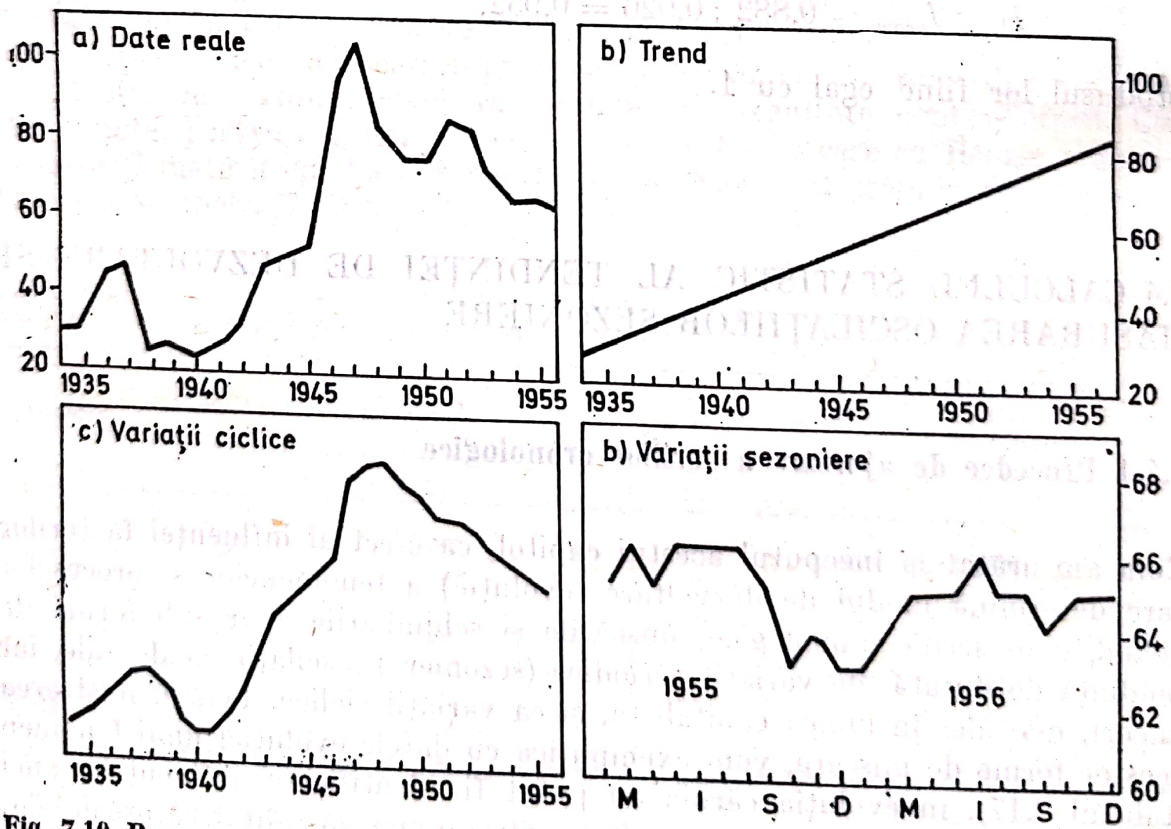


Fig. 7.10 Reprezentarea grafică a evoluției prețului grîului în comerțul internațional, în anii 1934—1956.

Curba din graficul *a* reprezintă datele empirice, respectiv nivelurile medii anuale ale prețului grîului. Din desfășurarea ei în cursul anilor putem desprinde acțiunea factorilor ciclici, întîmplători și sezonieri. În pofida acestor influențe, linia de bază (fundamentală) a mișcării prețurilor, de-a lungul celor 20 de ani, arată o creștere continuă (graficul *b*). Oscilația ciclică (graficul *c*) are alura ciclului capitalist „clasic" (8—10 ani). Paralel cu factorul ciclic a acționat, în perioada 1940—1947, al doilea război mondial cu efectele sale, seceta din Europa, care a determinat creșterea prețurilor, ca urmare a cererii sporite. Influența factorilor incidentali a predominat din nou în 1950—1951, odată cu reînarmarea și agresiunea din Coreea. Destinderea intervenită în 1952 a contribuit, alături de stocurile din America și recoltele bune, la scăderea prețurilor. A urmat o perioadă de depresiune, cu unele excepții sezoniere. Oscilațiile sezoniere (graficul *d*) arată o oarecare regularitate, în sensul că prețurile apar mai urcate în trimestrele II și III (perioada de legătură între recolta veche și cea nouă) și mai coborîte în celelalte trimestre. Și această situație este sub influența unei diversități de factori; unii — recolta bună sau presiunea stocurilor — acționează în direcția scăderii prețurilor, alții, dimpotrivă (gerurile în timpul iernii compromit recoltele, îngreunează transporturile scumpind navlul). Asemenea influențe complexe se reflectă tocmai în oscilațiile din graficul *d*.

Foarte succinta analiză în acest exemplu invederează, pe de o parte, existența unor mișcări diferite în evoluția fenomenelor, pe de altă parte, scoate în evidență necesitatea și interesul de a le analiza separat. Așa a apărut necesitatea de a stabili tendința fundamentală a evoluției fenomenelor, pe o durată medie sau mai lungă de timp, de a calcula trendul lor, după expresia adoptată pe plan internațional și în acest scop de a efectua ajustarea sau netezirea seriilor cronologice, după anumite procedee (metode).

Prin *ajustare* se înțelege operația de înlocuire a valorilor empirice ale unei serii statistice cu valori calculate prin metode statistice-matematice, eliminîndu-se variațiile întîmplătoare și scoțîndu-se în evidență tendința esențială de dezvoltare. Schema după care se face ajustarea (aproximarea) valorilor trebuie să concorde cu tendința reală a fenomenului, stabilită prin analiză prealabilă. Dacă metoda de ajustare este bine aleasă, seria ajustată reușește să exprime, cu claritate, linia caracteristică, tipică, de evoluție a fenomenului în cadrul complexului cauzal și al condițiilor ce determină mișcarea.

Pentru ajustarea seriilor statistice se folosesc metoda grafică, metode mecanice și metode analitice. Acestea se aplică și în operațiile de interpolare și extrapolare.

Ca expresie statistică a liniei sau direcției fundamentale de mișcare, obținută prin ajustare, *trendul* este un mijloc de exprimare sintetică a unei dinamici medii, este o mărime medie dinamică. În ce privește forma grafică, el se poate înfățișa ca o linie ascendentă sau descendentă, dreaptă sau frîntă, ca o curbă exponențială, parabolică sau hiperbolică și chiar ca o orizontală. Mai uzuale sînt formele dreptei și ale parabolei, dar cel mai des folosită este linia dreaptă. În linii generale, trendurile pot fi liniare și neliniare.

Metoda de ajustare grafică. Cum arată și denumirea, această metodă se sprijină mai ales pe reprezentarea seriei statistice cronologice sub forma unei cronograme, histograme sau a altei forme adecvate și în *trasarea vizuală* a liniei de tendință, urmărind alura curbei datelor empirice pe rețeaua graficului.

Scanned with OKEN Scanner

O astfel de ajustare are, vădit, un caracter aproximativ, conținînd impreciziile inerente însești metodei folosite. O aproximare mai bună cere nu numai o oarecare îndemînare, ci și o intuire a tendinței, pentru care se cere experiență. Metoda grafică este un procedeu elementar de ajustare, orientativ și poate fi folosit ca o operație preliminară, servind ca indicație în alegerea funcției matematice care să corespundă mai bine sensului tendinței.

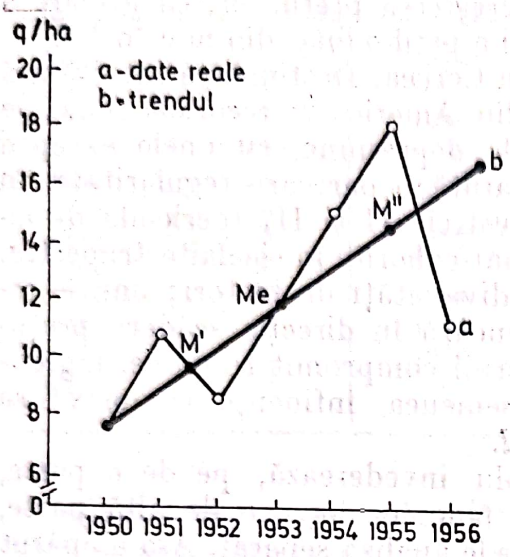


Fig. 7.11 Trendul creșterii producției medii de porumb în anii 1950—1956, trasat prin procedeul medianei [121].

Alt procedeu elementar de trasare grafică este cel bazat pe utilizarea medianei și a subseriilor, constînd în trasarea pe grafic a unei linii drepte (curbe), care să se apropie cît mai mult de toate punctele ce reprezintă valorile reale (empirice) ale seriei. În acest scop se determină punctul care are ca abscisă mijlocul șirului de variație, iar ca ordonată valoarea mediei aritmetice a seriei. Dreapta de ajustare trebuie să treacă prin punctul M, corespunzător medianei Me. Întrucît prin punctul Me pot trece mai multe drepte, procedeul se îmbunătățește împărțind seria în două părți egale și determinînd cîte un punct M pentru fiecare parte, astfel cum se vede în fig. 7.11 [121].

În acest exemplu, locul medianei a fost calculat astfel:

$$Me = \frac{n+1}{2} = \frac{7+1}{2} = 4,$$ corespunzător anului 1953, astfel că $Me =$ $= 11{,}2$ q/ha, valoarea medie a seriei fiind

$$\bar{y} = \frac{y}{n} = \frac{81{,}9}{7} = 11{,}7 \text{ q/ha.}$$

Punctul Me se determină cu abscisa corespunzătoare anului 1953 și cu ordonata corespunzătoare valorii de 11,7. Punctele M' și M'' se fixează cu ajutorul valorii medii de 9,7 q/ha, pentru prima parte a seriei și a valorii medii de 14,6 q/ha, pentru partea a doua.

Metode de ajustare mecanică. Pentru o mai bună aproximare a valorilor se folosesc procedee care constau în aplicarea mecanică și succesivă a unor formule alese în prealabil pentru toți termenii seriei care se ajustează, precum metoda sporului mediu, metoda ritmului mediu și metoda mediilor mobile.

Scanned with OKEN Scanner

Cea mai ușoară dintre acestea se consideră *metoda sporului mediu*, care constă într-un calcul simplu al sporului mediu anual din sporul absolut, pentru o serie întreagă și aceasta este apoi ajustată prin însumarea sporului mediu la nivelul ajustat al anului precedent, pe baza formulei

$$y_n = y_0 + n\bar{\Delta},$$

în care

$\bar{\Delta}$ este sporul mediu.
y — termenii seriei.

Aplicînd formule la exemplul concretizat în seria din tabelul 7.18 sporul absolut total este dat de diferența 4930 — 2113 = 2817, iar sporul absolut mediu: $\frac{2\,817}{6} = 469{,}5$.

Tabelul 7.18

INDICATORI AI PRODUCȚIEI DE ENERGIE ELECTRICĂ ÎN ȚARA X, ÎN PERIOADA 1968—1974

Ani	Nivel anual al producției, în mil. kWh	Spor absolut (în lanț)	Ritm de creștere, în % (în lanț)	Nivel anual ajustat al producției, mil. kWh
1968	2113	—	—	2113,0
1969	2473	360	117,0	2582,5
1970	2886	413	116,7	3052,0
1971	3411	525	118,2	3521,5
1972	3697	286	108,4	3991,9
1973	4340	643	117,3	4460,5
1974	4930	590	113,6	4930,0
	—	2817	—	—

Dacă transpunem în grafic datele din ultima coloană a tabelului, obținem o linie dreaptă (fig. 7.12). Deficiența esențială a acestei metode constă în faptul că tendința este determinată de primul și ultimul termen al seriei, influența termenilor intermediari fiind complet anihilată.

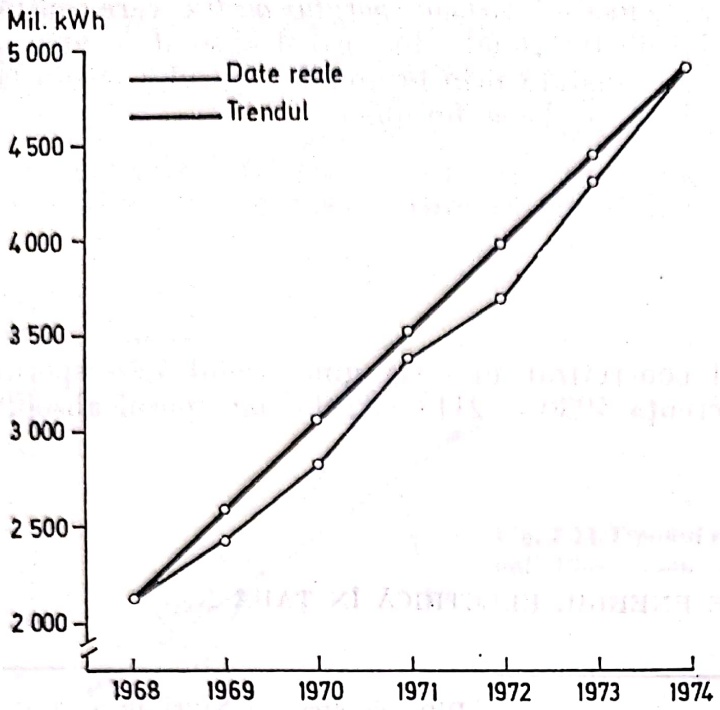
Metoda ritmului mediu conduce la rezultate aproximativ similare ca variație și tendință generală. Aici aproximarea termenilor se operează pe baza relației existente între termenul inițial al seriei (y_0) și ritmul mediu (r), aplicîndu-se formula

$$y_n = y_0 \cdot r^n,$$

în care r este ridicat la puterea egală cu numărul termenilor anteriori. Valoarea r este determinată ca medie geometrică sau, mai general, ca funcție exponențială.

Se poate trage concluzia generală că cele două metode mecanice prezentate dau rezultate mulțumitoare numai dacă există o regularitate de variație, absolută (progresie aritmetică) sau relativă (progresie geometrică) în cadrul seriei ajustate.

Scanned with OKEN Scanner



7.12 Trendul producției de energie electrică în țara X, în anii 1968—1974 (determinat prin metoda sporului mediu).
Întocmit pe baza datelor din *Anuarul statistic al R. S. România 1965 și 1972.*

Ajustarea prin metoda mediei mobile. Deși se bazează tot pe un calcul simplu, această metodă reprezintă un procedeu mai avansat, constituind o trecere către metodele analitice de calculare a trendului. Ea se bazează pe principiul compensației și constă în înlocuirea termenilor empirici ai seriei cu termeni calculați pe cîteva intervale. Calculîndu-se o medie din termeni corespunzători unui număr constant de perioade, abaterile termenilor reali ai seriei de la mediile mobile, care li se substituie prin seria ajustată, se compensează între ele. În felul acesta se obține o linie de dezvoltare mai mult sau mai puțin ondulată, în funcție de numărul termenilor din care s-a calculat media și de abaterile în serie.

Media mobilă poate fi calculată pentru trei sau mai mulți termeni, dar se recomandă ca numărul lor să nu fie mai mare de 9, oricîți termeni ar cuprinde seria reală. De asemenea, se cere ca perioadele pentru care se calculează media să fie egale. Dacă alegem, de exemplu trei termeni, media se va calcula astfel:

$$\bar{y}_I = \frac{y_1 + y_2 + y_3}{3}\ ;\quad \bar{y}_{II} = \frac{y_2 + y_3 + y_4}{3}\ ;\quad \bar{y}_{III} = \frac{y_3 + y_4 + y_5}{3}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Prin urmare, fiecare medie din trei termeni va cuprinde două elemente vechi și unul nou, astfel că media calculată se deplasează de-a lungul seriei, de unde și numele de medie mobilă. Dacă numărul termenilor incluși în medie este impar $(2n + 1)$, fiecare medie mobilă va înlocui termenul situat la mijlocul perioadei supuse ajustării. Dacă în calcul se include un număr par de termeni (2n), media se va plasa în punctul $n + \frac{1}{2}$ de pe intervalul luat în calcul. Metoda mediilor mobile se aplică îndeosebi la ajustarea seriilor

Scanned with OKEN Scanner

care prezintă oscilaţii mari, de obicei cu caracter sezonier. Pentru stabilirea numărului de termeni ai mediei, după o analiză prealabilă a fenomenului, se vor alege atîţia termeni cîţi corespund mai bine perioadelor de schimbări în evoluţia fenomenului, astfel că este necesar să se ţină seamă totdeauna de mărimea seriei, frecvenţa şi amplitudinea oscilaţiilor, precum şi de alţi factori care influenţează evoluţia.

Pentru exemplificarea calculului mediei mobile folosim datele lunare pe trei ani ale evoluţiei producţiei de energie electrică din tabelul 7.18, pentru care formăm grupe de cinci şi şapte termeni; seria lunară fiind mare obţinem astfel reprezentarea grafică din fig. 7.13. S-a considerat că perioada mai

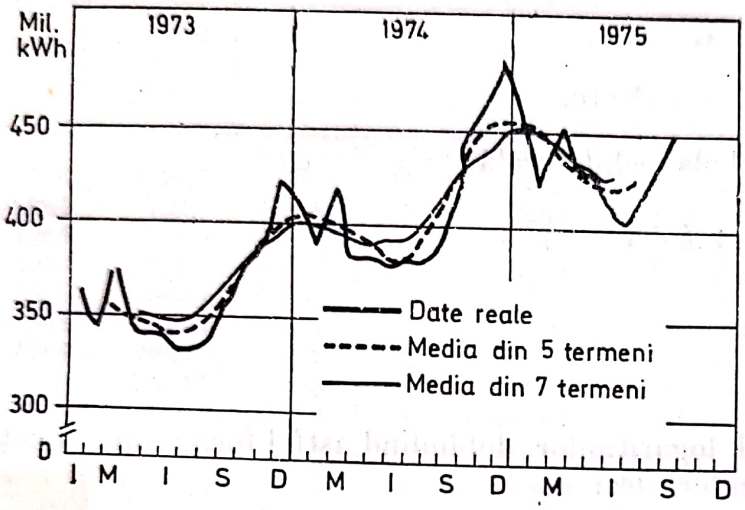


Fig. 7.13 Ajustarea datelor producţiei de energie electrică în ţara X, în lunile anilor 1973 — 1975, pe baza mediei mobile.

potrivită pentru aplicarea metodei în cazul de faţă este cea de 6—7 luni. „Netezirea" seriei reale în modul arătat scoate în evidenţă tendinţa de creştere a fenomenului.

Metoda mediei mobile este frecvent utilizată de statisticieni şi adesea o găsim aplicată în calculele de tendinţă a fenomenelor pe plan internaţional. Nu trebuie însă să uităm că această metodă are un caracter empiric şi cu ajutorul ei nu se pot formula legităţi. De aceea, pentru o exactitate mai mare, se utilizează metoda celor mai mici pătrate.

Aplicarea metodei celor mai mici pătrate. Întrucît caracteristicile acestei metode au fost arătate cu ocazia tratării ritmului mediu de creştere, vom reaminti numai că prin aplicarea ei se urmăreşte găsirea ecuaţiei liniei drepte sau curbe care să înlocuiască datele reale ale seriei cronologice, satisfăcînd următorul postulat: suma pătratelor diferenţei $(y - Y)$ dintre y, adică mărimile reale ale termenilor seriei şi Y, care simbolizează mărimile obţinute prin calcul, respectiv $\Sigma(y - Y)^2$, trebuie să fie minimă. Tipul ecuaţiei (funcţiei) urmează să fie ales în concordanţă cu natura fenomenului cercetat, cu datele reale ale seriei cronologice şi cu acţiunea diferiţilor factori, se înţelege pe baza analizei prealabile. Valorile ajustate, care trebuie să exprime cît mai obiectiv linia de mişcare a fenomenului, se numesc în mod frecvent ecuaţii de estimaţie, ecuaţii de regresie şi ele oglindesc trendul seriei.

Scanned with OKEN Scanner

Tipurile de ecuaţii (funcţii) folosesc polinoame de diferite grade. Pentru o dreaptă liniară, ecuaţia de gradul 1 cu formula:

$$Y = a + bx,$$

pentru polinomul de gradul 2,

$$Y = a + bx + cx^2,$$

pentru polinomul de gradul 3,

$$Y = a + bx + cx^2 + dx^3,$$

pentru polinomul de gradul 4,

$$Y = a + bx + cx^2 + dx^3 + ex^4 \text{ etc.}$$

De asemenea, pentru hiperbola echilaterală

$$Y = a + \frac{b}{x} \text{ sau } Y = a + b\frac{1}{x},$$

pentru funcţia exponenţială

$$Y = a \cdot b^x,$$

care se rezolvă cu ajutorul logaritmilor, dobîndind astfel forma unei funcţii liniare a logaritmilor termenilor, deci

$$\log Y = \log a + x \log b.$$

În metodologia ajustării seriilor dinamice există şi alte curbe folosite, precum curba logistică, curba lui Gompertz, apoi curba cumulativă numită şi integrală sau în clopot, care manifestă o analogie cu cea logistică.

Ajustarea presupune deci folosirea metodei celor mai mici pătrate. De pildă, în cazul polinomului de gradul 2

$$Y = a + bx + cx^2$$

sistemul de ecuaţii normale va fi

$$\text{I} \quad \Sigma y = Na + b\Sigma x + c\Sigma x^2$$

$$\text{II} \quad \Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\text{III} \quad \Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4.$$

Dacă se ia o origine arbitrară în seria cronologică, sistemul de ecuaţii normale se simplifică:

$$\text{I} \quad \Sigma y = Na + c\Sigma x^2$$

$$\text{II} \quad \Sigma xy = b\Sigma x^2$$

$$\text{III} \quad \Sigma x^2y = a\Sigma x^2 + c\Sigma x^4.$$

Scanned with OKEN Scanner

Pentru exemplificarea aplicării funcției liniare și a parabolei, folosim seria cronologică din tabelul 7.19, care ne dă posibilitatea să apreciem asupra cărei funcții să ne fixăm alegerea, în concordanță cu evoluția fenomenului. Întrucît seria este formată de un număr impar de ani, originea arbitrară, 0, o formează în acest exemplu anul de la mijlocul seriei, adică 1967, calculul simplificîndu-se în modul arătat.

Tabelul 7.19

DATELE PENTRU AJUSTAREA SERIEI CRONOLOGICE PRIVIND EXPORTUL DE ENERGIE ELECTRICĂ DIN ROMÂNIA, ÎN ANII 1963–1971

Ani	Date reale, mil. kWh y	x	x^2	xy	xy^2	$y = a + bx$	$Y = a + bx + cx^2$
1963	21,7	−4	16	−86,8	337,2	−35,6	143,2
1964	122,7	−3	9	−368,1	1104,3	377,0	422,8
1965	577,0	−2	4	−1154,0	2308,0	789,6	738,0
1966	1235,2	−1	1	−1235,0	1235,2	1202,2	1083,4
1967	2069,1	0	0	0	0	1614,8	1486,8
1968	2243,1	1	1	2243,1	2243,1	2027,4	1880,2
1969	2683,5	2	2	5367,0	10744,0	2440,0	2234,6
1970	2333,8	3	9	7001,4	21004,2	2852,6	2550,2
1971	3247,2	4	16	12988,8	51956,2	3265,2	2829,4
	14533,3	0	60	24756,4	90924,2	14533,2	13368,6

După *Anuarul statistic al R. S. România 1972*

Calculul parametrilor a și b, descris cu ocazia prezentării metodei celor mai mici pătrate, ne dă în acest exemplu următoarele valori: pentru funcția liniară: $a = \frac{\Sigma y}{n} = 1614{,}8$; $b = \frac{\Sigma xy}{x^2} = 412{,}6$; pentru funcția parabolică: $a = 1416{,}8$; $b = 412{,}6$.

Determinarea diferenței $\Sigma (y - Y)^2$ și minimului ei, ca postulat al metodei, ne arată că *trendul parabolic* dă o mai bună aproximare a valorilor ajustate, corespunzînd mai bine caracteristicilor evoluției seriei cronologice reale. Reprezentarea grafică a seriei reale și a trendurilor calculate cu ajutorul celor două funcții o dă fig. 7.14.

Cînd seria cronologică reală prezintă oscilații ca în figura 7.15 [121], trendul calculat cu aplicarea curbei parabolice descrie mai întîi o evoluție descendentă și apoi una ascendentă. În asemenea cazuri se recomandă folosirea unui număr mai mare de termeni ai seriei, pentru a se vedea dacă curba descrisă de datele reale este repetabilă sau numai întîmplătoare. Analiza capătă o mai mare precizie, dacă seria se împarte în două părți și ajustarea se face separat pentru fiecare din ele. Acest procedeu devine necesar atunci cînd prin ajustare se urmărește obținerea unui suport rațional pentru calcule de prognoză.

Scanned with OKEN Scanner

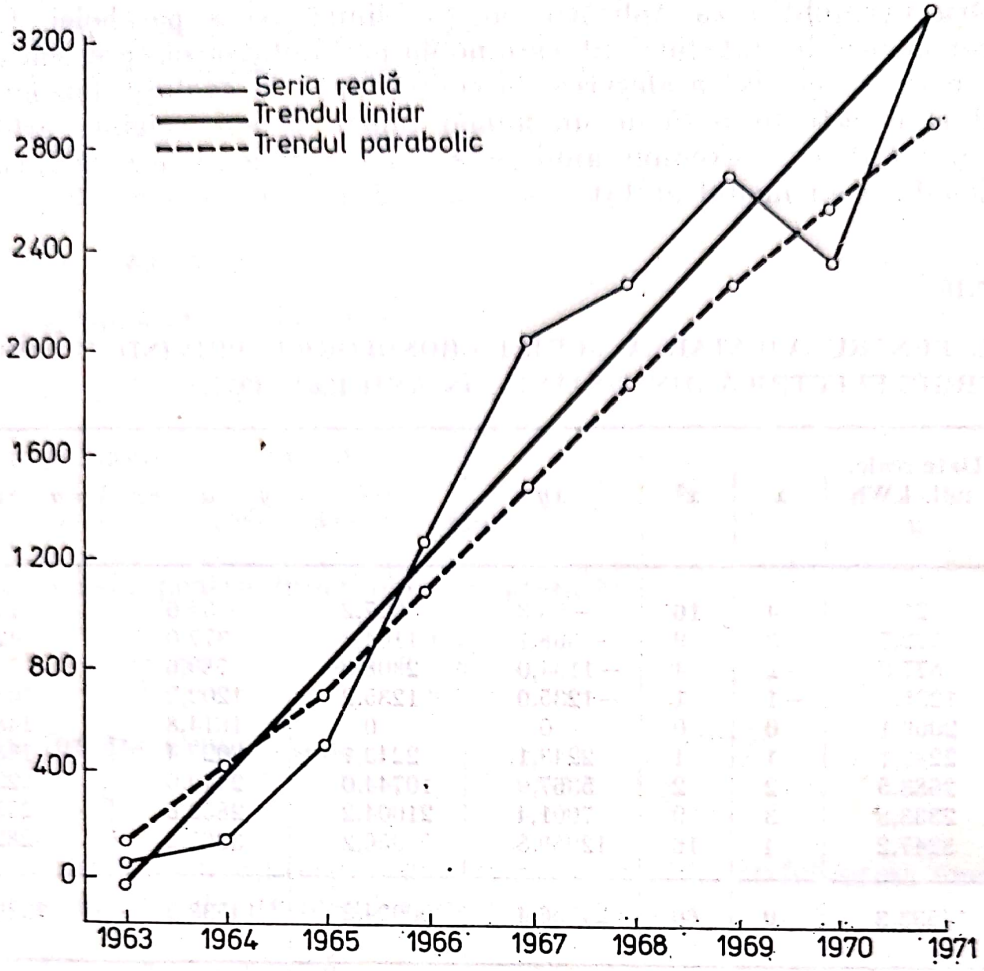


Fig. 7.14 Ajustarea seriei cronologice privind exportul de energie electrică din România, în anii 1963—1971, prin determinarea trendului liniar şi parabolic.

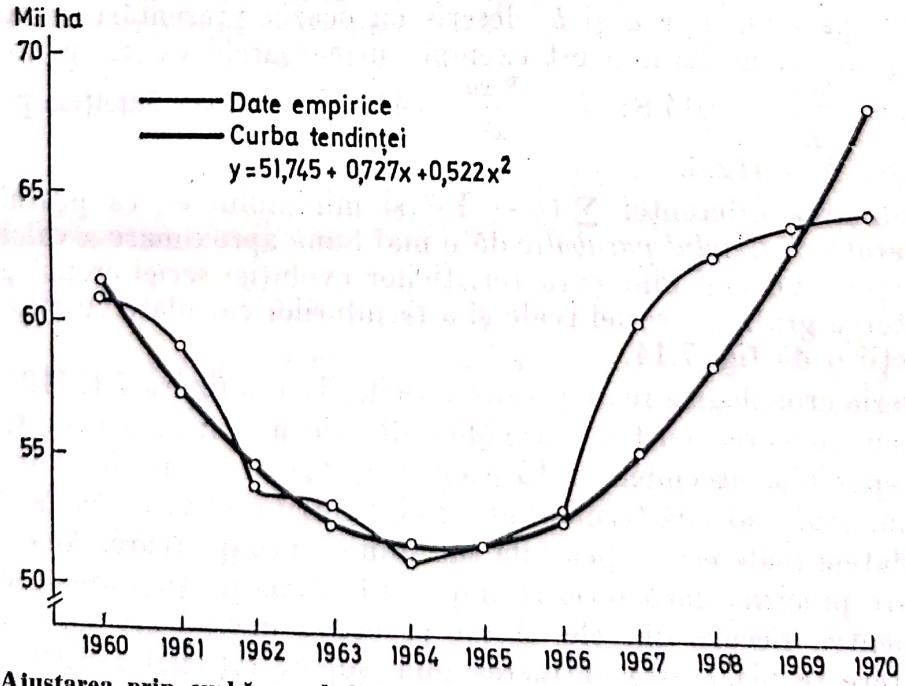


Fig. 7.15 Ajustarea prin curbă parabolică a seriei cronologice privind suprafeţele cultivate cu plante textile, în România, în anii 1960—1970 [121].

Scanned with OKEN Scanner

Pentru ilustrarea altui tip de curbă, folosim datele din tabelul 7.20, în care este urmărită evoluţia preţului de cost în funcţie de creşterea producţiei. Un astfel de calcul al tendinţei poate fi aplicat cu succes în planificare şi el este efectuat cu ajutorul *hiperbolei*, întrucît în acest caz nu se caută o mărime absolută, ci o unitate de legătură.[30]

Tabelul 7.20

PRODUCŢIA DE AER COMPRIMAT A UNEI ÎNTREPRINDERI ÎN CURSUL UNUI AN ŞI CALCULUL COSTULUI AFERENT VOLUMULUI PRODUCŢIEI

Lunile	Producţia 1000 m.c. (x)	Costul, DM/1000 mc. (w)	$\frac{1}{x}$ (10^{-3})	$\frac{1}{x^2}$ (10^{-6})	$\frac{w}{x}$ (10^{-3})
Ianuarie	1795	11,59	0,557103	0,310364	6,456824
Februarie	1650	11,89	0,606061	0,367309	7,206060
Martie	1750	10,57	0,571429	0,326531	6,154285
Aprilie	1595	11,43	0,626959	0,393078	7,166144
Mai	1850	12,67	0,540540	0,292184	6,848648
Iunie	2050	10,18	0,487805	0,237954	4,965853
Iulie	2200	10,03	0,454545	0,206612	4,559090
August	2446	9,61	0,408831	0,167143	3,928863
Septembrie	2338	10,14	0,427716	0,182941	4,337040
Octombrie	2157	10,53	0,463607	0,214931	4,789058
Noiembrie	2025	11,48	0,493827	0,243865	5,669135
Decembrie	2020	9,59	0,495050	0,245074	4,747525
		129,71	6,133473	3,187986	66,828525

Ecuaţia funcţiei unei hiperbole, aplicată în acest caz, este

$$Y = \frac{a}{x} + b,$$

în care

a este nivelul mediu al costului producţiei,

b — cheltuielile necesare pentru producerea a 1000 mc aer comprimat.

În calcul avem nevoie de două ecuaţii, şi anume:

$$\Sigma_w = a\left(\Sigma \frac{1}{x}\right) + nb \quad \text{şi} \quad \Sigma \frac{w}{x} = a\left(\Sigma \frac{1}{x^2}\right) + b\left(\Sigma \frac{1}{x}\right).$$

[30] A fost utilizat exemplul elaborat de ED. WÜNSCHE şi K. H. BEIER, *Die Trendberechnung, ein Hilfsmittel der Planung*, „Der Industriebetrieb", Berlin (R. D. Germană), nr. 1 şi 7/1957.

Scanned with OKEN Scanner

Pe baza datelor din tabelul 7.20 vom avea:

$$a = \frac{\left(\Sigma w \cdot \Sigma \frac{1}{x}\right) - \left(n \cdot \Sigma \frac{w}{x}\right)}{\left(\Sigma \frac{1}{x}\right)^2 - \left(n \cdot \Sigma \frac{1}{x^2}\right)} = \frac{(129{,}71 \cdot 6{,}133473 \cdot 10^{-3}) - (12 \cdot 66{,}828525 \cdot 10^{-3})}{6{,}133473 \cdot 10^{-3})^2 - (12 \cdot 3{,}187986 \cdot 10^{-6})} =$$

$$= \frac{(795{,}572783 \cdot 10^{-3}) - (801{,}942300 \cdot 10^{-3})}{(37{,}619491 \cdot 10^{-6}) - (38{,}255832 \cdot 10^{-6})} = \frac{-6369{,}517 \cdot 10^{-6}}{-0{,}637 \cdot 10^{-6}} = 10014{,}96$$

$$b = \frac{\left(\Sigma \frac{w}{x} \cdot \Sigma \frac{1}{x}\right) - \left(\Sigma w \cdot \Sigma \frac{1}{x^2}\right)}{\left(\Sigma \frac{1}{x}\right)^2 - \left(n \cdot \Sigma \frac{1}{x^2}\right)} =$$

$$= \frac{(66{,}828525 \cdot 10^{-3} \cdot 6{,}133473 \cdot 10^{-3}) - (129{,}71 \cdot 3{,}187986 \cdot 10^{-6})}{(6{,}133473 \cdot 10^{-3})^2 - (12 \cdot 3{,}187986 \cdot 10^{-6})} =$$

$$= \frac{(409{,}880954 \cdot 10^{-6}) - (431{,}513664 \cdot 10^{-6})}{(37{,}619491 \cdot 10^{-6}) - (38{,}255832 \cdot 10^{-6})} = \frac{-3{,}622710 \cdot 10^{-6}}{-0{,}636341 \cdot 10^{-6}} = 5{,}69.$$

Introducînd rezultatele în ecuația funcției hiperbolice,

$$y = \frac{a}{x} + b = \frac{10\,014{,}96}{x} + 5{,}69$$

vom putea determina cheltuielile posibile, necesitate de producția din fiecare lună (x), care — cum se vede mai jos — sînt așezate în ordinea creșterii producției:

x	1595	1650	1750	1795	1850	2020	2025	2050	2157	2200	2338	2446
$\frac{a}{x}$	6,28	6,07	5,72	5,58	5,41	4,96	4,95	4,89	4,64	4,55	4,28	4,09
$+ b =$	5,69	5,69	5,69	5,69	5,69	5,69	5,69	5,69	5,69	5,69	5,69	5,69
$= Y$	11,97	11,76	11,41	11,27	11,10	10,65	10,64	10,58	10,33	10,24	9,97	9,78

Hiperbola, care în grafic (fig. 7.16) apare ca o curbă transcendentă, ne ajută să stabilim în care luni producția a fost realizată peste sau sub nivelul mediu al costului.

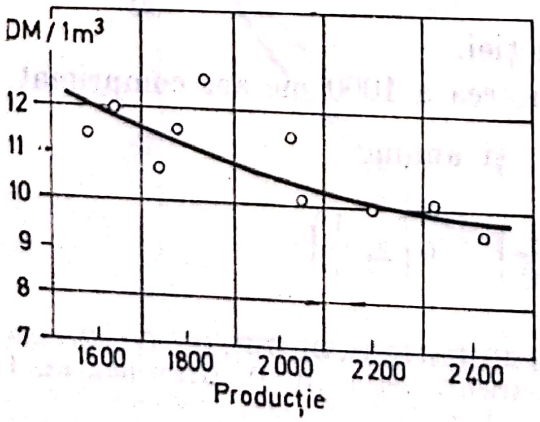


Fig. 7.16 Curba reducerii costului aferent producției de aer comprimat, determinată cu ajutorul ecuației hiperbolei (după K. H. Beier).

Scanned with OKEN Scanner

Observarea atentă a fenomenelor social-economice, mai ales atunci cînd prin calculele de tendinţă se urmăreşte fundamentarea previziunilor şi în cazurile în care seriile empirice oferă o evoluţie pronunţat crescătoare, ne conduce la concluzia că o asemenea creştere „la infinit" este greu de conceput şi că de la o anumită limită, numită *punct de inflexiune*, ascensiunea curbei va fi frînată, fenomenul intrînd într-un proces de autofrînare, care s-ar putea continua chiar cu o evoluţie coborîtoare.

Aceasta ne conduce la aşa-numitele curbe de creştere asimptotică. Dacă dreapta $y = a + bx$ descrie un spor constant al creşterii sau descreşterii, iar curba exponenţială $y = ab^x$ implică un raport constant al schimbării şi deci un raport constant al schimbării modificării absolute, curbele asimptotice de creştere vor arăta un spor crescător al modificării în condiţiile unei rate descrescătoare a schimbării.

Curbele de acest gen sînt: curba exponenţială modificată, curba lui Gompertz şi curba logistică.

Curba exponenţială modificată descrie un trend în care sporul creşterii se diminuează cu un procent constant, iar curba se apropie de o limită superioară, denumită asimptotă.

Ecuaţia curbei exponenţiale modificate este:

$$Y = k + ab^x,$$

în care k este asimptotă.

Curba logistică, denumită astăzi şi curba Pearl-Reed[31], are forma cea mai simplă:

$$\frac{1}{y} = k + ab^x.$$

În mod obişnuit ea se scrie:

$$y = \frac{k}{1 + 10^{a+bx}} \quad \text{sau} \quad \frac{k}{1 + e^{a+bx}},$$

în funcţie de folosirea logaritmilor zecimali sau naturali.

Rezolvarea se poate face cu diferite metode. Folosind metoda punctelor alese este necesar să luăm trei ani x_0, x_1 şi x_2, echidistanţi între ei: unul cam la începutul perioadei, al doilea, la mijlocul perioadei, al treilea, aproximativ spre sfîrşitul perioadei. Perioada de timp trebuie să fie suficient de mare.

Valorile ajustate sînt trei: y_0, y_1 şi y_2. Cele trei constante se obţin astfel:

$$k = \frac{\Sigma y_0 y_1 y_2 - y_1^2(y_0 + y_2)}{y_0 y_2 - y_1^2}$$

$$a = \log \frac{k - y_0}{y_0}$$

$$b = \frac{1}{n} \log \left[\frac{y_0(k - y_1)}{y_1(k - y_0)} \right].$$

[31] Folosită mai întîi în 1838, de matematicianul belgian P. F. VERLUST, curba logistică a fost redescoperită în 1920 de demografii americani R. PEARL şi L. J. REED *(On the matematical Theory of Population Growth*, Metron, 1923*)*, şi mai ales de RAYMOND PEARL, *(The Biology of Population Growth*, 1925*)*, iar în cercetări economice a fost aplicată de S. KUZNETS *(Secular Movement in Production and Prises*, 1830*)* şi de DUPRIEZ *(Des mouvements économiques généraux*, 1947*)*.

Scanned with OKEN Scanner

Aplicată inițial la ajustarea evoluției populației S.U.A., curba logistică a început să fie folosită în econometrie, mai ales la estimarea cererii de bunuri de folosință îndelungată, caracterizată printr-un punct de „saturație", precum și la alte fenomene și în diferite studii de marketing. Teoria modernă a sistemelor, mai ales a sistemelor biologice, a reactualizat importanța curbei logistice.

Alura trendului calculat cu ajutorul curbei logistice (fig. 7.17) a determinat denumirea de „creștere în S", punctul M corespunzînd maximului

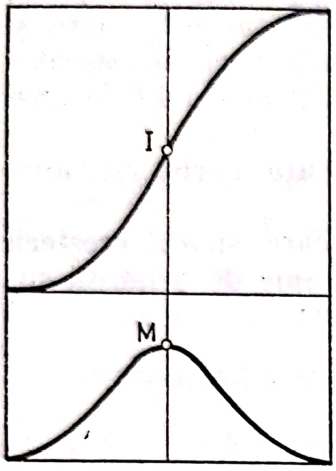


Fig. 7.17 Curba logistică și în formă de clopot]51].

curbei normale sau integrale în formă de clopot. Jan Tinbergen, care a atras atenția asupra acestei similitudini, a făcut chiar afirmația că viteza reacției, caracterizată de o lege logistică, se repartizează aproape corespunzător legii normale (curba Laplace-Gauss) [51].

Asemănătoare cu curba logistică este *curba Gompertz*[32], care descrie un trend, cu sporurile de creștere a logaritmilor micșorîndu-se cu un procent constant. Valorile naturale ale trendului vor arăta o rată descrescătoare a creșterii, dar rata nu descrește nici cu un spor constant și nici cu un procent constant. Formula este

$$Y = ka^{b^x},$$

folosită în operația de ajustare ia forma logaritmică

$$\log Y = \log k + (\log a)b^x.$$

Calculul se face aproximativ după aceeași schemă ca la curba exponențială modificată. Într-o serie cronologică împărțită în trei subperioade, formula este

$$b^n = \frac{\Sigma_3 \log Y - \Sigma_2 \log Y}{\Sigma_2 \log Y - \Sigma_1 \log Y}$$

$$\log a = (\Sigma_2 \log Y - \Sigma_1 \log Y)\frac{b-1}{(b^n-1)^2}$$

$$\log k = \frac{1}{n}\left[\Sigma_1 \log Y - \left(\frac{b^n-1}{b-1}\right)\log a\right],$$

în care n reprezintă numărul termenilor (anilor) din fiecare subperioadă.

[32] Formulată pentru prima dată în anul 1825, de actuarul englez Benjamin Gompertz (1779—1865) în memoriul *On the nature of the function expresive of the law of human mortality*, curba lui Gompertz are larg cîmp de aplicare în demografie, la ajustarea și extrapolarea probabilităților de deces și supraviețuire la vîrstele înaintate etc.

Scanned with OKEN Scanner

Dacă se urmărește calculul lui k fără să se determine în prealabil log a și b, atunci formula folosită este:

$$\log k = \frac{1}{n}\left[\frac{(\Sigma_1 \log Y)(\Sigma_3 \log Y) - (\Sigma_2 \log Y)^2}{\Sigma_1 \log Y + \Sigma_3 \log Y - 2\Sigma_2 \log Y}\right].$$

Curba Gompertz se folosește în econometrie, în modelele creșterii economice, în marketing etc.

7.4.2 Calculul variațiilor sezoniere

Observarea modului în care variază anumite fenomene naturale sau social-economice scoate în evidență existența unei periodicități, care se repetă cu o anumită consecvență, de pildă variația lunară de la an la an a precipitațiilor atmosferice, a numărului de călători pe mijloacele de transport în comun, în diferite orașe, în cursul zilei, a numărului de căsătorii contractate în raport cu anotimpul etc. Să urmărim datele care reflectă un astfel de fenomen în tabelul 7.21.

Reprezentarea grafică (fig. 7.18) a datelor din tabelul 7.21 înfățișează curba evoluției fenomenului în cursul celor zece ani cu variațiile sale lunare,

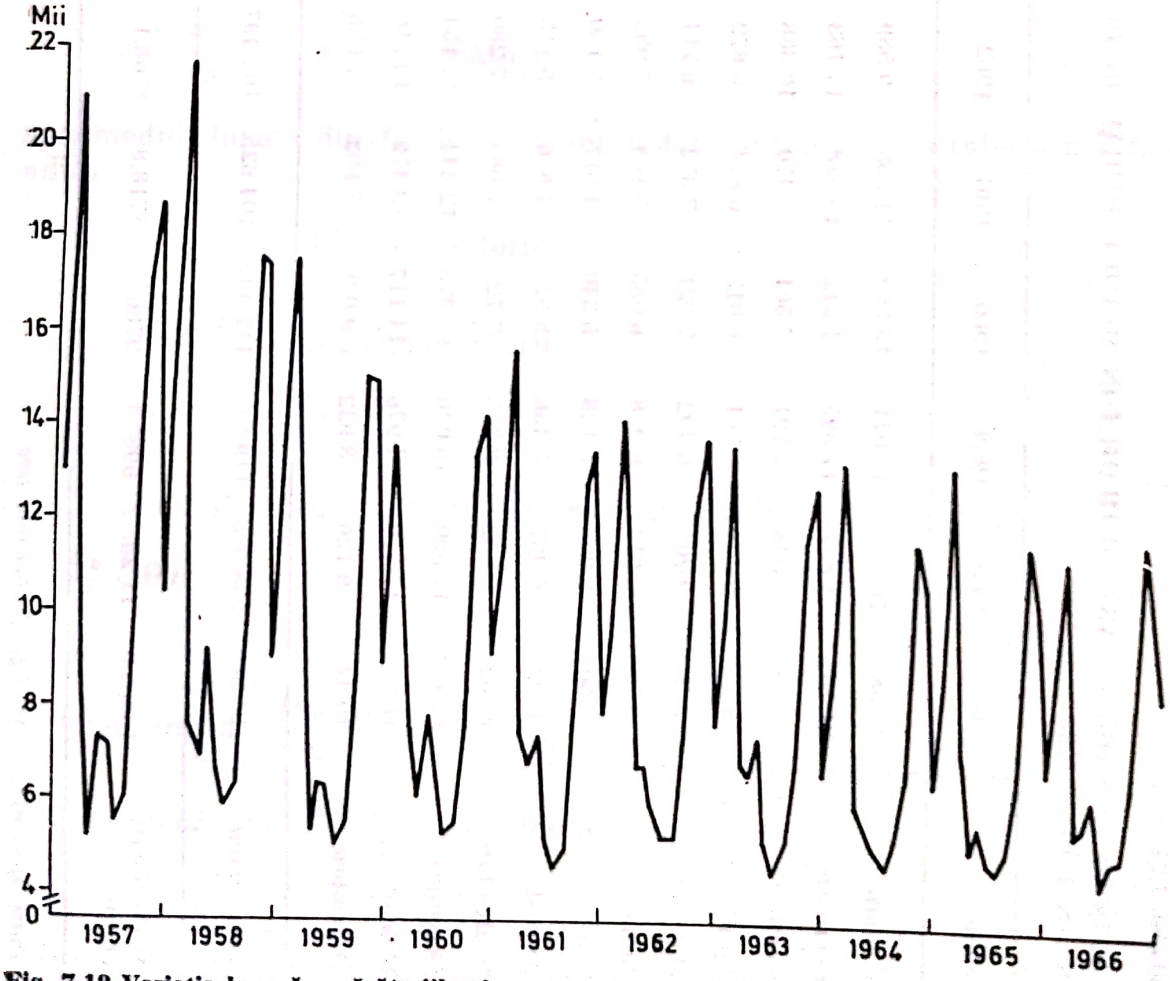


Fig. 7.18 Variația lunară a căsătoriilor în mediul rural din România, în anii 1957—1966.

Tabelul 7.21

VARIAȚIA LUNARĂ A CĂSĂTORIILOR DIN MEDIUL RURAL ÎN ROMÂNIA ÎN ANII 1957—1966

Lunile	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	Total lunar
Ianuarie	13 583	16 266	13 071	13 569	11 358	9 589	9 707	8 810	8 461	8 760	113 174
Februarie	21 167	21 867	17 395	21 212	15 528	14 163	13 533	13 294	13 111	11 155	162 425
Martie	8 750	7 538	12 791	7 561	7 470	10 386	6 902	10 401	7 037	5 421	84 257
Aprilie	5 335	6 908	5 361	6 013	6 824	6 672	6 595	5 865	4 954	5 564	60 091
Mai	7 458	9 205	6 492	7 821	7 372	6 714	7 226	5 355	5 464	6 114	69 221
Iunie	6 705	6 612	6 238	6 085	5 169	5 953	5 137	5 029	4 827	4 387	56 142
Iulie	5 504	5 908	5 138	5 340	4 617	5 130	4 464	5 689	4 625	4 872	50 287
August	6 025	6 367	5 456	5 522	4 876	5 115	5 116	5 296	4 947	4 996	53 716
Septembrie	10 627	10 162	8 775	7 723	7 685	7 280	6 814	6 474	6 401	6 540	78 681
Octombrie	17 124	17 590	15 020	13 373	12 444	12 154	11 403	11 455	11 527	11 596	133 686
Noiembrie	18 842	17 519	14 976	14 117	13 479	13 716	12 685	10 613	11 043	10 197	137 187
Decembrie	10 442	9 126	8 932	9 096	7 803	7 625	6 771	6 313	6 844	8 286	81 238
Total anual	131 562	135 068	119 845	117 432	104 625	104 497	96 353	93 594	89 241	87 888	1 080 105
Medii lunare pe an	10963,5	11 255,7	9987,1	9786,0	8718,8	8708,1	8029,4	7799,5	7436,8	7324,0	—

După *Anuarul statistic al R. S. România 1970.*

Scanned with OKEN Scanner

care se repetă cu destulă regularitate în aceleaşi luni ale anilor. Maximile din perioada de iarnă şi linia de tendinţă ce o putem trasa cu ochiul, lent coborîtoare, ca şi restrîngerea consecventă a amplitudinii oscilaţiilor de-a lungul trendului, se datoresc acţiunii convergente a mai multor factori, pe care analiza îi poate scoate în evidenţă. Se observă clar că mişcarea reală a fenomenului rezultă din influenţa conexă a două funcţii: una sezonieră $S(t)$ şi una extrasezonieră $E(t)$, care include tendinţa şi eventual alt tip de mişcare. Conexiunea celor două funcţii este privită în statistică prin prisma a două forme: 1) forma aditivă, adică $S(t) + E(t)$, în care variaţia amplitudinii este independentă de $E(t)$ şi apare mai constantă şi 2) forma multiplicativă, adică $E(t).[1 + S(t)]$, caracterizată printr-o proporţionalitate a celor două curbe şi crescătoare în timp. Realitatea prezintă mai frecvent mişcări de prima formă [51].

Metodologia actuală determină variaţiile sezoniere prin coeficienţi calculaţi după diverse metode, din care alegem pentru exemplificare pe cele mai des aplicate.

Metoda rapoartelor la media generală. Pentru ilustrarea acestei metode urmează calculul, pentru care se folosesc aici datele din tabelul 7.21. Se calculează mai întîi media generală pe cele 120 de luni date, utilizîndu-se ultima cifră din rîndul penultim al tabelului, adică

$$\bar{X} = \frac{1\,080\,105}{120} = 9000{,}9 \text{ căsătorii,}$$

apoi mediile lunare din datele pe 10 ani (notate în coloana 1 a tabelului 7.22), adică

$$\bar{x}_I = \frac{113\,174}{10} = 11317{,}4 \text{ căsătorii}$$

$$\bar{x}_{II} = \frac{162\,425}{10} = 16242{,}5 \text{ căsătorii}$$

. . . . . .

. . . . . .

Coeficientul sezonier pentru fiecare lună rezultă din raportul

$$\frac{\bar{x}_n}{\bar{X}} \cdot 100$$

astfel că pentru

luna ianuarie $\frac{11\,317{,}4}{9\,000{,}9} \cdot 100 = 125{,}7\,\%$

luna februarie $\frac{16242{,}5}{9000{,}9} \cdot 100 = 180{,}5\,\%$

. . . . . .

. . . . . .

Coeficienţii obţinuţi demonstrează că, în ce priveşte căsătoriile contractate în mediul rural din ţara noastră, proporţia cea mai ridicată în ordinea

Scanned with OKEN Scanner

mărimii revine lunilor februarie, noiembrie, octombrie şi ianuarie, respectiv în medie cu 80,5%, 52,4%, 48,5% şi 25,7% mai multe căsătorii faţă de media generală lunară [88]. Caracterul sezonier s-a redus însă mult în ultimele două decenii, faţă de perioada interbelică.[33]

Metoda rapoartelor la tendinţă. Această metodă se foloseşte atunci cînd seria datelor lunare este ajustată analitic, pe baza unei funcţii. Coeficienţii sezonieri pot fi calculaţi ca rapoarte între mediile lunare şi datele lunare ajustate.

Folosind aceleaşi date din tabelul 7.21, pentru aplicarea acestei metode, operaţiile se desfăşoară după cum urmează [88].

Mai întîi se ajustează mediile lunare, notate în rîndul ultim din tabelul 7.21 prin metoda celor mai mici pătrate, de pildă cu ajutorul ecuaţiei dreptei; pe baza formulei cunoscute:

$$10\ a + 55\ b = 90008{,}9$$

$$55\ a + 385\ b = 457195{,}8$$

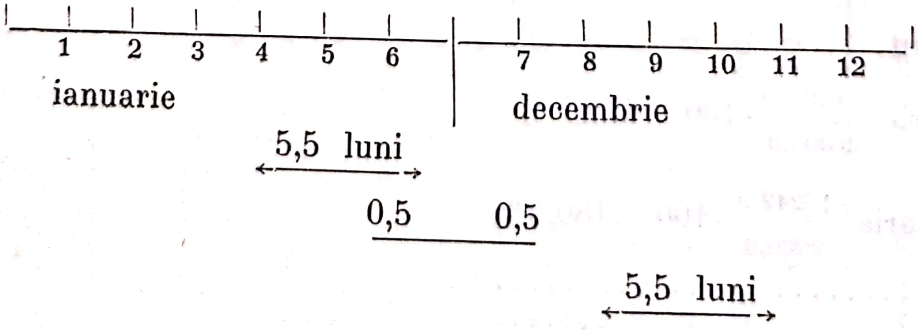
$$a = 11524{,}4; \qquad b = -458{,}8,$$

iar pentru efectuarea calculului ne servim de următorul tabel ajutător:

Ani	Medii lunare x	t	t^2	$\bar{x}t$
1957	10963,5	1	1	10963,5
1958	11255,7	2	4	22511,4
1959	9987,1	3	9	29961,3
1960	9786,0	4	16	39144,0
1961	8718,8	5	25	43594,0
1962	8708,1	6	36	52248,6
1963	8029,4	7	49	56205,8
1964	7799,5	8	64	62396,0
1965	7436,8	9	81	66931,2
1966	7324,0	10	100	73240,0
	90008,9	55	385	457195,8

Scăderea medie lunară $\frac{b}{12} = \frac{458{,}8}{12} = 38{,}23$ căsătorii.

Conform schemei



[33] C. BRĂDESCU, *Oscilaţii periodice în frecvenţa încheierii căsătoriei*, „Revista de statistică" nr. 12/1964.

cele 12 medii lunare ajustate se obţin astfel:

ianuarie $9000{,}9 + (-38{,}23) \cdot (-5{,}5) = 9211{,}1$
februarie $9000{,}9 + (-38{,}23) \cdot (-4{,}5) = 9172{,}9$
. . . . . .
. . . . . .
iunie $9000{,}9 + (-38{,}23) \cdot (-0{,}5) = 9020{,}0$
iulie $9000{,}9 + (-38{,}23) \cdot (\ \ 0{,}5) = 8981{,}8$
. . . . . .
. . . . . .
decembrie $9000{,}9 + (-38{,}5) \cdot (\ \ 5{,}5) = 8790{,}6$

În continuare, se raportează media fiecărei luni la media corespunzătoare ajustată şi astfel se obţin coeficienţii sezonieri, notaţi în tabelul 7.22 (coloana 5).

Tabelul 7.22

MEDIILE LUNARE EMPIRICE ŞI AJUSTATE, PRECUM ŞI COEFICIENŢII SEZONIERI AI CĂSĂTORIILOR DIN MEDIUL RURAL AL ROMÂNIEI, PE ANII 1957—1966

Luni	Medii lunare pe 10 ani (x)	Abateri de la media generală lunară ($x - X$)	Medii lunare ajustate (x')	Coeficienţi sezonieri		Rangul coeficienţilor din:	
				metoda A $\frac{\bar{x}}{\bar{X}} \cdot 100$	metoda B $\frac{\bar{x}}{\bar{x}'} \cdot 100$	Col. 4	Col. 5
0	1	2	3	4	5	6	7
Ianuarie	11317,4	+2316,5	9211,1	125,7	122,9	4	4
Februarie	16242,5	+7241,6	9172,9	180,5	177,1	1	1
Martie	8425,7	− 575,2	9134,7	93,6	92,2	5	6
Aprilie	6009,1	−2991,8	9096,5	66,8	66,1	9	9
Mai	6692,1	−2078,8	9058,2	76,9	76,4	8	8
Iunie	5614,2	−3386,7	9020,0	62,4	62,2	10	10
Iulie	5028,7	−3972,2	8981,8	55,9	56,0	12	12
August	5371,6	−3629,3	8943,6	59,7	60,1	11	11
Septembrie	7868,1	−1132,8	8905,3	87,4	88,4	7	7
Octombrie	13368,7	+4367,8	8867,1	148,5	150,8	3	3
Noiembrie	13718,7	+4717,8	8828,9	152,4	155,4	2	2
Decembrie	8123,8	− 877,1	8790,6	90,3	92,4	6	5

În funcţie de condiţiile concrete în care evoluează variaţiile sezoniere, au fost aduse unele amendamente metodelor prezentate. De pildă, considerînd cazurile în care influenţează trendul secular sau mişcarea ciclică, W. M. Pearsons a pus la punct un calcul mai complicat, cunoscut sub denumirea *metoda lanţurilor de rapoarte*, folosind mediana în locul mediei aritmetice, pentru a evita abaterile extreme. Mediana este considerată ca reprezentînd raportul lunar dintre elementul lunar de ordinul k şi elementul lunar precedent. În ce priveşte metoda B, tratată aici, desigur că pot fi folosite atîtea procedee de determinare a coeficienţilor sezonieri, cîte procedee de calculare a trendului pot fi aplicate. De pildă, Simon Kuznets a propus să se folosească media mobilă a celor 12 luni, pe considerentul că oscilaţiile sezoniere se desfăşoară în această perioadă [51].

Calculele efectuate după cele două metode aplicate de noi, plecînd de la aceleaşi date empirice, ne demonstrează — cum se constată în ultimul tabel —

că diferențele dintre coeficienții aceleiași luni sînt relativ mici și că există o ierarhizare aproape identică a lunilor în raport cu mărimea coeficienților. Este posibil ca la diverse fenomene mici variații de la lună la lună să fie influențate și de inegalitatea numărului de zile ale fiecărei luni. Făcînd această remarcă, N. Georgescu-Roegen propune [47] ca atunci cînd se studiază izolat variațiile sezoniere să se lucreze și cu medii zilnice, nu cu date globale, luîndu-se în considerare numărul zilelor *utile*, adică cu medii zilnice în cazul fenomenelor meteorologice și demografice, iar la fenomenele economice cu medii rezultate din împărțirea datelor lunare la numărul zilelor lucrătoare din luna respectivă.

Plecînd de la astfel de constatări, ca și de la diverse aspecte noi ale fenomenelor conjuncturale ale pieței capitaliste în lumea contemporană, în ultimul deceniu s-au făcut încercări de a găsi un nou mod de abordare a seriilor cronologice și nu putem încheia acest capitol fără a atrage cel puțin atenția asupra lui.

Analiza spectrală a seriilor cronologice. Cu toate că seriile cronologice au prezentat de foarte mult timp un interes deosebit pentru cunoașterea fenomenelor și proceselor social-economice, metodele de analiză au evoluat încet. Spre deosebire de diverse domenii din cîmpul științelor naturii, unde observarea oferă serii de mărimi cu caracter de continuitate, seriile cronologice cu care ne-am ocupat în acest capitol au un caracter discret, adică observațiile pe care ele le reflectă emană din anumite momente, de obicei echidistante. Între aceste momente pot apărea evenimente care nu ies în evidență în datele înregistrate. În analiza unei serii se disting de obicei trei etape: 1) investigarea fenomenului; 2) elaborarea unui model pentru cunoașterea lui și 3) aplicarea (valorificarea) modelului. La rîndul său, construirea modelului presupune, de asemenea, conform precizării lui P. Whithe, parcurgerea a trei etape: a) limitarea alegerii la un număr finit de tipuri plauzibile pentru modelare; b) selecționarea unuia din aceste tipuri și c) precizarea valorii numerice a parametrilor în tipul de model ales. Cu timpul s-a constatat însă că adecvarea modelului la datele disponibile nu constituie soluționarea problemei esențiale a analizei, mai cu seamă cînd se studiază relațiile dintre serii, care sînt atît de complicate, încît este foarte greu de găsit modelul potrivit, care să fie și ușor inteligibil. Dealtfel, statisticienii s-au convins de timpuriu că majoritatea seriilor reale nu sînt simple înșiruiri de date independente [34] și au încercat să scoată în evidență componentele unei serii cronologice, pentru a le studia separat prin disociere.

Metodele folosite în acest scop s-au dovedit prea simple față de dificultatea problemei și, încă în 1936, A. Wald a avansat unele propuneri, cu ocazia studiului său asupra variațiilor sezoniere, dar ele au rămas necunoscute. În mod logic, s-au găsit cercetători care s-au adresat metodelor folosite în științele naturii, în special celor aplicate în fizică și astronomie. [35]

Analiza spectrală, definită în fizică drept cercetare a naturii chimice a unei substanțe, prin examinarea spectrului pe care-l formează, pe baza faptului că poziția liniilor și a benzilor de emisiune și de absorbție ale unui spectru este caracteristică fiecărei substanțe [36], a oferit denumirea pentru noua metodă, prin prelungirea în științele naturii a conceptului de analiză formulat de

[34] C. W. J. GRANGER, *Analyse spectrale des séries temporelles en économie*, Dunod, Paris, 1969.
[35] O. MORGENSTERN, în prefață la lucrarea citată a lui C.W.J. Granger.
[36] *Dicționar politehnic*, Editura tehnică, București, 1957.

Scanned with OKEN Scanner

Joseph Fourier[37], concomitent cu progresul contemporan al matematicii și tehnicii de calcul.

Prin fundamentarea cercetării seriilor cronologice pe principiile analizei spectrale, astfel cum ele au fost dezvoltate în ultimele decenii, ne găsim în fața posibilității unei revizuiri a vechii metodologii de analiză a acestor serii, care apar mult mai complicate decît se cunoștea pînă acum. Se obțin astfel informații noi cu privire la metodele de eliminare a oscilațiilor anuale și mijloace noi de construire a unor filtre eficiente.[38]

Analiza spectrală, ca metodă de cercetare a seriilor cronologice, constituie o generalizare a teoriei descompunerii formulate de J. Fourier și ea își propune să studieze proprietățile intrinseci ale seriilor și amplitudinile componentelor periodice: de asemenea, prin analiza *încrucișată* ea pune în evidență legăturile dintre serii, măsurînd decalajele dintre ele. Plecînd de la definirea *spectrului* ca „diagramă de descriere a fiecărei frecvențe a unei serii cronologice", cunoașterea lui oferă informații cu privire la proprietățile seriei, scoțînd în evidență componentele sale importante. Această metodă își valorifică însă fertilitatea mai ales în analiza unui ansamblu de serii.

În fizică, analiza spectrală s-a putut dezvolta ușor (mai cu seamă în optică și telecomunicații) datorită caracterului staționar al seriilor, ceea ce a permis simplificări. Aplicată în economie, ea s-a lovit de dificultăți, întrucît aici fenomenele se caracterizează prin nestaționaritate; de aceea, teoria spectrală curentă, aplicată în economie, nu este nici pe departe completă, dar s-au conturat unele tehnici aplicabile.[39] Un concept important pe care-l poate valorifica analiza spectrală îl constituie conexiunea inversă (feed-back) dintre variabilele economice.

Cercetările efectuate în țările capitaliste s-au concentrat mai mult asupra fenomenelor ciclice conjuncturale, precum și asupra principiului accelerării în materie de stocuri. S-a putut face constatarea că în economie sînt frecvente cazurile în care spectrul unei serii economice descrește, pe măsură ce frecvența crește. Între principalele probleme care se cer adîncite sînt considerate, în primul rînd, extinderea cercetărilor asupra seriilor cronologice nestaționare și generalizarea teoriei cu privire la conexiunea inversă, problema intervalelor de încredere cu privire la spectrele estimate; mai sînt de înlăturat anumite rezerve referitoare la procedeele folosite în ce privește eliminarea componentei sezoniere. Metodologia analizei spectrale reclamă și este susceptibilă încă de importante îmbunătățiri, dar — cum precizează O. Morgenstern — ea va contribui la fundamentarea cercetării economice pe concepte moderne și încercate.

[37] *J. Fourier* (1768—1830), matematician francez, descoperitorul seriilor trigonometrice și a integralei ce-i poartă numele, care s-au dovedit de o importanță considerabilă pentru analiza mișcărilor vibratorii, a fenomenelor ondulatorii și sonore, precum și în studierea analitică a propagării căldurii (lucrarea sa fundamentală: *Théorie analytique de la chaleur*, 1822).

[38] Un pas hotărîtor în aplicarea teoriei spectrale în analiza fenomenelor social-economice l-a marcat *Proiectul asupra seriilor cronologice*, alcătuit în cadrul „Programului de cercetări econometrice" de la Universitatea din Princeton, sub direcția lui Oskar Morgenstern, consilier fiind John Tukey și colaboratori Michael Godfrey, Michio Hatanaka, Hermann Karreman, Mitsuo Luzuki, Thomas Wonnacott și Cliwe W. J. Granger, autorul volumului *Analyse spectrale des séries temporelles en économie*. În proiectul asupra seriilor temporale au fost dezvoltate principiile metodologice fundamentale ale analizei spectrale în economie.

[39] O aplicație instructivă: EUGEN PECICAN, *Analiza spectrală și descompunerea seriilor dinamice*, în „Revista de statistică" nr. 8/1973, unde, după unele considerente teoretice de ordin general, precum și cu privire la aplicarea metodei la seriile cronologice din economie, modelul reprezintă însumarea componentelor: trend, sezonalitate constantă și variabilă reziduală.

Scanned with OKEN Scanner

CAPITOLUL 8
Corelație

8.1 CONSIDERAȚII PRELIMINARE

În cercetarea statistică intervin adesea repartiții în care fiecărei unități a populației considerate îi corespund simultan modalitățile a două sau a mai multor caracteristici, de aceeași natură sau de natură diferită. Asemenea repartiții se impun ori de cîte ori știm sau suspectăm o relație între caracteristici și dorim să o precizăm; studiul lor suscită un interes capital. De multă vreme, de exemplu, greutatea excesivă a fost incriminată ca reprezentînd un factor de influență asupra speranței de viață. Îndelungi fapte de observație, transcrise statistic sub formă de repartiții bidimensionale și minuțios analizate, au confirmat ipoteza emisă și au pus în evidență mărimea riscului de scurtare a anilor de viață în grupele de populație cu exces de greutate. Pe o cale similară sînt investigate astăzi corelațiile posibile între diverși parametri ai vieții moderne și bolile de etiologie multifactorială și, în măsura în care ipotezele se verifică, cercetătorii se angajează cu îndrăzneală în estimații numerice. Cine se recunoaște bunăoară în tipul de comportament sisific (mereu agitat) trebuie să ia aminte — avertizează rezultatele cercetărilor recente — că el duce de 2,3 ori mai mult la angină pectorală, față de ceilalți, și de 4,3 ori mai frecvent la infarct.

Construirea repartițiilor statistice cu mai multe caracteristici nu comportă vreo dificultate deosebită. Regulile elementare de claritate și ordine, formulate în legătură cu repartițiile după o singură caracteristică, prezidează și la construirea acestor repartiții. De asemenea, aceleași distincții există în privința naturii caracteristicilor și modalității lor de variație: unitățile de observare pot fi grupate, concomitent, fie după caracteristici calitative, fie după caracteristici cantitative, fie după ambele tipuri de caracteristici în combinare, caracteristicile cantitative fiind, la rîndul lor, fie cu variație discretă, fie cu variație continuă.

Să considerăm o populație descrisă simultan după două variabile x și y. Notînd prin $x_1, x_2, \dots, x_i \dots, x_n$ valorile observate ale variabilei x, prin $y_1, y_2, \dots, y_j, \dots, y_m$ valorile observate ale variabilei y, iar prin f_{ij} numărul

unităților populației, care au valoarea x_i a variabilei x și totodată valoarea y_j a variabilei y, repartiția îmbracă forma:

Tabelul 8.1

y \ x	x_1	x_2	...	x_i	...	x_n	*Total*
y_1	f_{11}	f_{21}	...	f_{i1}	...	f_{n1}	$f_{\cdot 1}$
y_2	f_{12}	f_{22}	...	f_{i2}	...	f_{n2}	$f_{\cdot 2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_j	f_{1j}	f_{2j}	...	f_{ij}	...	f_{nj}	$f_{\cdot j}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_m	f_{1m}	f_{2m}	...	f_{im}	...	f_{nm}	$f_{\cdot m}$
Total	$f_{1\cdot}$	$f_{2\cdot}$	...	$f_{i\cdot}$	...	$f_{n\cdot}$	$f_{\cdot\cdot}$

Tabelul alcătuit poartă numele de *tabel cu dublă intrare*. Cînd variabilele sînt continue, coloanele ca și rîndurile se asociază la intervale și se convine să se desemneze prin x_i și y_j centrele intervalelor. Deoarece numărul total de observații este $f_{\cdot\cdot}$, se scrie:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} = f_{\cdot\cdot} \tag{8.1}$$

Expresia (8.1) se poate descompune fie efectuînd suma mai întîi în raport cu i, apoi în raport cu j, fie în raport cu j, apoi în raport cu i. Evident, în ambele cazuri se totalizează frecvențele parțiale și se obține suma tuturor frecvențelor ($f_{\cdot\cdot}$):

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} = \sum_{j=1}^{m}\left(\sum_{i=1}^{n} f_{ij}\right) = \sum_{j=1}^{m}(f_{1j} + f_{2j} + \dots + f_{nj})$$

și suma din urmă dezvoltată:

$$(f_{11} + f_{12} + \dots + f_{1m}) + (f_{21} + f_{22} + \dots + f_{2m}) + \dots + (f_{n1} + f_{n2} + \\ + \dots + f_{nm}).$$

La fel:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij} = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} f_{ij}\right) = \sum_{i=1}^{n} f_{i1} + f_{i2} + \dots + f_{im}$$

și desfășurat:

$$(f_{11} + f_{21} + \dots + f_{n1}) + (f_{12} + f_{22} + \dots + f_{n2}) + \dots + (f_{1m} + f_{2m} + \\ + \dots + f_{nm})$$

Tabelul 8.2

y \ x	Sub 20	20—24	25—29	30—34	35—39	40—44	45—49
Sub 20	6 584	44 833	12 738	1 311	179	26	11
20—24	1 870	44 278	23 956	4 862	1 141	215	69
25—29	112	3 222	6 671	3 788	1 543	433	147
30—34	21	367	1 051	1 672	1 463	673	261
35—39	3	105	299	658	1 016	880	435
40—44	—	28	78	179	439	616	594
45—49	1	10	9	37	129	255	477
50—54	—	2	2	7	22	49	182
55—59	—	7	6	1	1	10	28
60 și peste	1	4	6	1	—	2	13
f_x	8 592	92 856	44 816	12 516	5 933	3 159	2 217
xf_x	146 064	2 042 832	1 210 032	400 512	219 521	132 678	104 199
x^2f_x	2 483 088	44 942 304	32 670 864	12 816 384	8 122 277	5 572 476	4 897 353
$\sum yxf_{xy}$	2 668 728	40 508 534	26 268 489	10 211 744	6 656 892	4 706 436	4 199 638

Dacă se fac totalurile pe coloane ($f_{1\cdot}$, $f_{2\cdot}$, ... , $f_{n\cdot}$) și totalurile pe rînduri ($f_{\cdot 1}$, $f_{\cdot 2}$, ... , $f_{\cdot m}$), se găsesc, la marginile orizontale și verticale ale tabelului cu dublă intrare, două repartiții simple, numite *repartiții marginale*: repartiția integrală a variabilei x și repartiția integrală a variabilei y în populația totală. Altfel spus, sînt repartiții în funcție de o singură variabilă, făcîndu-se abstracție de cealaltă.

Exemplul 1. Tabelul 8.2 ilustrează corelația dintre vîrsta soțului (x) și vîrsta soției (y) în căsătoriile din R. S. România, în anul 1974.

Examinînd tabelul, observăm că pentru fiecare mărime a vîrstei soțului (x), avem o repartiție a vîrstei soției (y). O astfel de repartiție se numește *repartiție condiționată* (legată), în opoziție cu repartițiile marginale, care sînt repartiții „libere". Repartițiilor condiționate le asociem *medii condiționate* și abateri pătratice medii corespunzătoare. Dacă pentru fiecare repartiție y calculăm valoarea ei medie, obținem o serie de medii condiționate de valorile pe care le ia variabila x.

Fiecare medie condiționată se calculează ca medie aritmetică ponderată, după formula:

$$\hat{y}_x = \frac{\sum y f_{ij}}{\sum f_{ij}}, \tag{8.2}$$

în care

$\hat{y}_x$ este media condiționată;
y — valoarea mijlocului intervalului repartiției y;
f_{ij} — frecvențele perechilor de valori y, x.

Astfel, media condiționată pentru intervalul „sub 20 ani" este egală cu:

$$\hat{y}_x = \frac{(17\times 6584) + (22\times 1870) + (27\times 112) + (32\times 21) + (37\times 3) + (47\times 1) + (62\times 1)}{8\,592} =$$

$$= 18{,}27 \text{ ani.}$$

Scanned with OKEN Scanner

50—54	55—59	60 și peste	f_y	yf_y	y^2f_y	Σxyf_{xy}
5	3	4	65 694	1 116 798	18 985 566	25 381 731
25	9	12	76 437	1 681 614	36 995 508	41 037 678
46	7	7	15 976	431 352	11 646 504	12 407 364
74	21	26	5 629	180 128	5 764 096	6 132 256
179	51	60	3 686	136 382	5 046 134	5 269 614
336	104	181	2 555	107 310	4 507 020	4 750 410
461	226	380	1 985	93 295	4 384 865	4 699 060
283	270	696	1 513	78 676	4 091 152	4 420 312
59	125	507	744	42 408	2 417 256	2 493 636
28	90	1 132	1 277	79 174	4 908 788	4 821 368
1 496	906	3 005	175 496	3 947 137	98 746 889	111 413 429
77 792	51 642	186 310	4 571 582			
4 045 184	2 943 594	11 551 220	130 044 744			
3 456 024	2 549 724	10 187 220	111 413 429			

Se pot calcula, în exemplul luat, mediile lui x condiționate de valorile y. Seriile de medii condiționate și abaterile pătratice medii sînt următoarele:

Tabelul 8.3

Grupa de vîrstă a soțului	Vîrsta medie a soției	Abaterea pătratică medie	Grupa de vîrstă a soției	Vîrsta medie a soțului	Abaterea pătratică medie
Sub 20	18,27	2,48	Sub 20	22,73	3,15
20—24	19,83	3,04	20—24	24,40	3,84
25—29	21,71	3,98	25—29	28,76	5,54
30—34	25,50	5,61	30—34	34,04	6,96
35—39	30,32	6,92	35—39	38,64	7,81
40—44	35,47	7,31	40—44	44,27	8,59
45—49	40,30	7,94	45—49	50,37	8,26
50—54	44,43	7,79	50—54	56,18	6,81
55—59	49,37	8,00	55—59	58,80	6,54
60 și peste	54,68	7,65	60 și peste	60,90	4,33

Repartițiile empirice cu mai multe caracteristici au drept corespondent, în domeniul probabilităților, repartițiile de variabile aleatoare multidimensionale, cu noțiuni, foarte înrudite, de legi marginale, legi condiționate și chiar indicatori de corelație.

8.2 CONCEPTUL DE CORELAȚIE

Problema nouă și de mare importanță, pusă de studiul repartițiilor statistice multidimensionale, privește legăturile între variabilele observate și, prin ele, între fenomenele pe care le reprezintă.

Scanned with OKEN Scanner

Două fenomene se pot afla într-o legătură strictă, rigidă, exprimată prin formula $y = f(x)$ — pentru o valoare determinată a variabilei independente x, variabila dependentă y ia de asemenea o valoare determinată, discretă. Situaţia opusă este aceea a independenţei complete a două fenomene. Între aceste două extreme se întinde cîmpul larg al legăturilor suple, neunivoce, *legături de tip statistic* sau *stocastic* [1]: fiecărei valori x îi corespunde nu o singură valoare y, ci o „repartiţie legată" de valori y (legată de valoarea lui x).

Legătura statistică este generată de factori multipli, cu influenţe în sensuri şi de intensităţi diferite. Ea este expresia coacţiunii necesităţii şi întîmplării, necesitatea realizîndu-se prin intermediul jocului factorilor aleatori. Ca urmare, dependenţa nu se manifestă în fiecare caz în parte, ci numai în general şi în medie, într-un număr mare de cazuri.

Ideea de legătură între variabilele observate în colectivităţi statistice, dezvăluită şi transpusă în limbaj matematic de Şcoala biometrică Galton-Pearson, s-a constituit într-o teorie, numită *teoria corelaţiei statistice* sau *corelaţie* [2].

Corelaţia include două probleme fundamentale: prima constă în a descrie *legea de variaţie medie* a unei variabile în funcţie de o altă (sau alte) variabilă, cunoscută sub denumirea de *problema regresiei* şi rezolvată pe baza funcţiei de regresie, iar a doua, în caracterizarea intensităţii legăturii printr-un coeficient numeric, independent de unităţile de măsură ale variabilelor corelate [3].

Chiar de la începuturile ei, corelaţia se traducea matematicește prin „... formarea ecuaţiei ce dă valoare unei variabile în funcţie de o alta sau de mai multe altele" şi prin determinarea valorii coeficientului de corelaţie, care „ ... măsoară exactitatea cu care ecuaţia de regresie se aplică" [133]. Ulterior,

[1] Denumire întîlnită, se pare pentru prima oară, la A.A. Ciuprov, cu semnificaţia de legătură între variabilele aleatoare [56, p. 147]. Statisticianul suedez C. Charlier a numit legi stocastice legile relevate de statistică [47, p. 6].

[2] Termenul „corelaţie" este împrumutat din ştiinţele naturii. Naturalistul francez Georges Cuvier (1769—1832) formulează principiul corelaţiei astfel: „Orice fiinţă înzestrată cu organe formează un ansamblu, un sistem unic şi închis, ale cărui părţi se leagă mutual şi concură la aceeaşi acţiune definitivă printr-o reacţie reciprocă. Nici una din aceste părţi nu poate să se schimbe fără ca celelalte să nu se schimbe; şi, prin urmare, fiecare din ele, luată separat, influenţează şi se leagă de toate celelalte" (citat după [45]). Charles Darwin a folosit expresia „variabilitate corelativă" înţelegînd, tot aşa, raporturile reciproce dintre diversele părţi ale organismului. F. Galton a transpus termenul „corelaţie" în statistica matematică cu semnificaţia de raporturi reciproce între caracteristicile părinţilor şi fiilor. În studiul *Correlations and their measurement...*, Galton scrie: „...Corelaţiile se observă peste tot unde variaţiile a două fenomene se datoresc în parte uneia şi aceleiaşi cauze comune" [81]. În sens larg, etimologic (lat. *correlatio* = relaţie cu), cuvîntul corelaţie exprimă „relaţia, legătura reciprocă între două sau mai multe lucruri sau fenomene" [45].

[3] O primă formulare şi aplicare oarecum rudimentară a corelaţiei se întîlneşte în memoriul lui F. Galton, *Regression toward mediocrity in hereditary stature* (1886), inclus în cartea *Natural inheritance* (1889). Cu ajutorul tabelului de corelaţie şi al seriilor de medii legate, Galton a examinat legătura dintre înălţimea părinţilor şi înălţimea fiilor, postulînd existenţa unei regresii către valoarea medie: din părinţi cu talie mai mare decît media grupului se nasc copii cu o talie inferioară lor şi invers. În mod corespunzător, linia reprezentînd legătura între una din variabile şi mediile legate a primit denumirea de „linie de regresie". De aici, s-au ivit două probleme: a) care este ecuaţia matematică a liniei de regresie? şi b) este tendinţa de regresie la fel de puternică pentru diferite grupuri? Răspunsul matematic a fost dat de Karl Pearson, prin indicarea ecuaţiei liniei drepte adecvată problemei originale de regresie a lui Galton şi prin introducerea celebrului său coeficient de corelaţie. Astfel, s-au născut cele două probleme fundamentale ale teoriei corelaţiei: problema regresiei şi problema măsurării gradului de dependenţă [72, p. 657] [102, p. 331—332].

Scanned with OKEN Scanner

conținutul teoriei corelației s-a îmbogățit odată cu elaborarea metodelor neparametrice de măsurare a legăturilor, aplicabile la caracteristicile calitative.

În terminologia statistică actuală, conceptul de corelație apare, uneori, cu un înțeles foarte general, acoperind orice legătură statistică — fie între variabile cantitative, fie între variabile calitative [94]. Alteori, noțiunea de corelație semnifică numai relațiile dintre variabilele cantitative [62].

Într-un sens și mai restrîns, corelația nu ar fi decît o măsură a gradului de legătură între variabile cantitative, respectiv coeficientul sau raportul de corelație [62]. Evident, în această ultimă accepție, importanța calculelor de corelație în cercetările statistice este mult diminuată. După părerea noastră, metoda corelației întrunește întregul complex de procedee de investigație a conexiunilor statistice, conducînd la expresii numerice sintetice: ecuații de regresie, coeficienți de regresie, coeficienți de corelație, coeficienți de determinare, erori de estimare etc. Cu toate că metoda corelației, în sensul restrîns al cuvîntului, și metoda regresiei diferă prin premisele lor, totuși ele se împletesc atît de strîns în analiza statistică de ansamblu a legăturii încît valoarea lor de cunoaștere poate fi apreciată numai judecîndu-le împreună, într-o sinteză metodologică.

8.3 PREMISELE CALCULELOR DE CORELAȚIE

De fapt, corelația este doar o reflectare a legăturilor cu caracter complex existente în mod obiectiv între fenomenele de masă și, ca atare, nu trebuie să se uite că o reflectare poate ușor să fie o aparență înșelătoare. Practica îndelungată a calculelor de corelație prescrie cîteva premise, de a căror respectare depinde în măsură hotărîtoare valoarea de cunoaștere a indicatorilor de corelație.

a) Legăturile statistice sînt de natură probabilistă și deci „imperfecte". Această particularitate face dificilă descifrarea cauzelor generatoare de anumite comportări ale fenomenelor investigate. Înfăptuirea necesității în multicorelație cu întîmplarea complică mecanismul relației cauză-efect: nu avem să determinăm o singură cauză, ci o structură cauzală, în care cauzele participă inegal la producerea efectului sau efectul este produs cînd de unele cauze, cînd de altele și, nu rareori, acțiunea unei cauze depinde ea însăși de altă cauză [4]. Oricum, țelul cercetării trebuie să fie depistarea adevăratelor cauze, deoarece numai cunoașterea acestora permite explicația științifică a proceselor de masă și, pe această bază, în practică — provocarea de efecte dorite.

[4] Faptul că aspectul de necesitate se realizează, în cazul legăturii statistice, prin intermediul jocului factorilor aleatori, naște întrebarea dacă o astfel de legătură implică sau nu un conținut cauzal, ceea ce, pe un plan mai general, constituie problema cauzalității statistice. În controversele actuale, cîștigă teren opinia după care cauzalitatea funcționează și în procesele stocastice. Relațiile specifice acestor procese nu exclud cauzalitatea, ci, dimpotrivă, confirmă existența unui nou tip de relații cauzale — *relații cauzale statistice* [128]. Este interesant de observat utilizarea conceptului de cauzalitate statistică (în formularea „... case of statistical causation...") încă în anul 1909 de G. Udny Yule, în memoriul său privind aplicațiile metodei corelației în statistica socială și economică, prezentat la sesiunea a XII-a a Institutului Internațional de Statistică [133].

Scanned with OKEN Scanner

Atunci cînd se caută relaţiile cauzale dintre fenomenele de masă, se pot întîlni diferite situaţii, care, de fapt, se reduc la următoarele: o legătură nemijlocită între fenomene, unilaterală sau cu reacţii inverse (interdependenţă), o covariaţie a fenomenelor datorată unor cauze comune, un simplu paralelism întîmplător în variaţia a două sau mai multe fenomene. În prima categorie intră, bunăoară, dependenţele economice: productivitatea muncii — calificarea forţei de muncă, productivitatea muncii — înzestrarea tehnică, cererea de mărfuri — veniturile băneşti etc. sau interdependenţele: venit naţional — acumulare, producţie — cerere, productivitate — remunerare etc. Sînt exemple de raporturi cauzale obiectiv existente, descoperite şi indicate precis de ştiinţa economică. Este cunoscut, de exemplu, şi raportul dintre productivitatea muncii şi preţul de cost, în sensul că, sub influenţa creşterii productivităţii muncii, în general, scade preţul de cost. Faţă de celelalte dependenţe menţionate, aici apare însă o deosebire, şi anume variaţia concomitentă a celor două fenomene poate fi privită ca efectul unor cauze comune: calificarea forţei de muncă, înzestrarea tehnică a muncii, mecanizarea şi automatizarea proceselor de producţie, organizarea muncii etc. Cercetarea analitică a legăturii dintre productivitatea muncii şi preţul de cost, cu ajutorul metodelor statistico-matematice, prezintă fără îndoială importanţă, întrucît productivitatea muncii sintetizează înseşi influenţele cauzelor primare ale variaţiei preţului de cost. Dar, din punctul de vedere al eficienţei practice, este necesară extinderea cercetării, pe cît posibil, la cauzele imediate.

În alte cazuri, două fenomene se modifică în acelaşi sens sau în sensuri opuse, cu o anumită regularitate, ca urmare a modificării unui al treilea fenomen, fără ca între primele două fenomene să existe vreo legătură. Astfel, examinarea datelor statistice relevă covariaţia cererii de mărfuri şi a economiilor băneşti ale populaţiei. Un calcul de corelaţie, să spunem, ar fi lipsit de sens, fiecare variabilă fiind legată de veniturile băneşti ale populaţiei.

Asemenea cazuri, nu prea rare şi nici atît de simple, atrag atenţia că *sînt necesare eforturi pentru a se distinge, în complexitatea fenomenelor economice, legăturile cauzale de legăturile formale.* Succesul inferenţei spre o relaţie cauzală printr-o relaţie statistică depistată empiric poate fi asigurat de cunoaşterea temeinică a domeniului economic abordat şi de experienţa personală.

b) Datorită complexităţii cauzale, formalizarea matematică a dependenţelor statistice este inevitabil legată de unele simplificări. Astfel, se includ în calcul — după caz — unul sau mai mulţi factori cu influenţă esenţială, iar restul factorilor, global (o variabilă aleatoare). Întrucît nu este posibilă epuizarea factorilor, rezolvarea unui model de corelaţie oferă aproximaţii ale valorilor funcţiei. Cu alte cuvinte, în cercetarea legăturilor statistice ne mulţumim să ştim cum se modifică, în medie, variabila explicată, ca urmare a modificărilor variabilei sau variabilelor explicative. [5] Firește, pe măsura *cuprinderii*

[5] Denumirile clasice ale variabilelor corelate sînt: variabilă dependentă şi variabilă independentă. Un cuplu de termeni folosit în econometrie face distincţie între variabila endogenă — variabilă ale cărei valori sînt deduse din rezolvarea modelului — şi variabila exogenă — variabilă substituită „din afară", cu ajutorul căreia se determină valorile variabilei endogene. Se folosesc şi denumirile de variabilă explicată şi variabilă explicativă. Dacă prima se suprapune cu noţiunea de variabilă endogenă, cea de-a doua excede sfera noţiunii de variabilă exogenă, cuprinzînd şi variabila endogenă „întîrziată" sau „avansată" în timp, utilizată pentru explicarea variabilei endogene.

unui număr mai mare de variabile explicative, creşte gradul de aproximaţie a valorilor funcţiei.

Alegerea principalilor factori nu este o operaţie simplă; ea este legată de impedimente metodologice, tehnice şi de altă natură. În ajutorul cercetătorului vin, pe de o parte, cunoştinţele, experienţa, intuiţia, precum şi literatura de specialitate cu preocupări similare, iar, pe de altă parte, metodele statisticii matematice de verificare a existenţei legăturii şi de evaluare cantitativă a importanţei fiecărui factor.

Indicaţii prealabile despre factorii posibili de introdus în model oferă analiza teoretică şi logică a corelaţiilor. Pe această cale, se pot separa, apriori, factorii care au sau nu influenţă asupra fenomenului modelat şi se poate imagina chiar ierarhizarea factorilor în raport cu gradul lor de influenţă. De mare folos pentru alcătuirea listei factorilor, mai ales în cazul dependenţelor deosebit de complexe, este investigarea opiniei specialiştilor în domeniul abordat.

Alegerea apriorică a factorilor se continuă cu *verificarea existenţei legăturii* pe bază de criterii obiective, elaborate de statistica matematică. De reţinut că dependenţele bine ştiute nu pot fi negate, chiar dacă nu se verifică statistic pentru anumite perioade, colecţii de fapte sau din cauza unor deformări ale datelor statistice. Dependenţele cele mai importante, de exemplu în economia naţională, au fost descoperite şi indicate precis de ştiinţa economică; alte dependenţe nu sînt încă stabilite sub raport teoretic. În asemenea cazuri, cercetările statistice pot semnala dependenţe existente, dar necunoscute, pe baza observării de masă a fenomenelor, într-o perioadă suficient de mare, asigurîndu-se condiţia repetării lor. Acolo unde statisticianul merge pe o cale neumblată, unde poate să enunţe doar unele ipoteze despre existenţa dependenţei, trebuie să intervină verificarea statistică a ipotezelor emise. Dar şi în astfel de cazuri, primatul îl deţine analiza calitativă.

Mai departe, investigaţiile converg spre estimarea gradului de influenţă al fiecărui factor pentru care s-a admis ipoteza legăturii, în scopul de a reţine *factorii cu influenţă semnificativă*. Un procedeu de estimare poate fi compararea variaţiei explicate prin factorul cercetat cu variaţia totală a variabilei dependente. În ce priveşte modalităţile de selecţionare, sînt posibile două variante: alegerea pe calea introducerii progresive a variabilelor în model şi alegerea pe calea eliminării succesive a variabilelor din model. Potrivit primei variante, examinarea influenţelor se limitează, iniţial, la două sau trei dintre cele mai însemnate variabile independente şi apoi, dacă se constată că nu dau o explicaţie satisfăcătoare (apreciată, de pildă, după mărimea variaţiei explicate), se introduc noi variabile, una cîte una. În a doua variantă, ecuaţia de regresie cuprinde de la început toate variabilele independente, alese provizoriu, eliminîndu-se treptat variabilele cu influenţă neesenţială. Practica recomandă această variantă, dat fiind avantajul reducerii volumului de calcule [6].

În procesul selecţionării variabilelor, indiferent de sistemul folosit, trebuie totdeauna *să se tindă spre un model cît se poate de „economic"*. Pentru unele

[6] Principalele sisteme de eliminare succesivă a variabilelor independente sînt expuse de V.S. Nemcinov [101], care a elaborat o schemă proprie de rezolvare a sistemului de ecuaţii prin polinoamele lui Cebîşev, cunoscută sub denumirea de schema N. Folosirea largă a schemei N, în rezolvarea problemelor de corelaţie multiplă, se datorează avantajului său: în cazul includerii sau excluderii variabilelor nu este necesar să se repete toate calculele.

cercetări de început, în care „entuziasmul matematic nu era temperat de bunul simţ comun", a fost caracteristică includerea în calculele de corelaţie a unui număr de factori mai mare decît numărul observaţiilor, conducînd la coeficienţi înalţi de corelaţie, care, negreşit, nu sînt chiar atît de semnificativi.

Este evident că restrîngerea numărului de factori trebuie să aibă loc numai cu respectarea strictă a principiilor analizei de corelaţie. Simplificarea neîntemeiată a modelelor de corelaţie, teama de funcţii complexe, care se impun în mod obiectiv, generează pericolul vulgarizării procesului. La construirea oricărui model, cercetătorul este ameninţat de două pericole foarte sugestiv caracterizate de cunoscutul matematician R. Bellman: „Capcana suprasimplificării şi mlaştina supracomplicării" (citat după [101]).

c) Cînd se formulează o corelaţie multiplă, se presupune dependenţa variabilei y de variabilele $x_1, x_2, ..., x_n$ şi independenţa reciprocă a variabilelor $x_1, x_2, ..., x_n$. Dacă premisa din urmă nu este satisfăcută, ne aflăm în prezenţa fenomenului denumit *multicoliniaritate*, căruia în aplicaţiile practice nu i se dă atenţie prea mare sau care pur şi simplu este ignorat.

Consecinţele încălcării premisei independenţei variabilelor explicative se traduc în erori de estimare, uneori stranii, manifestate prin schimbarea sensului însuşi al legăturii reale [7].

Există metode menite să verifice existenţa multicoliniarităţii şi să corecteze efectele ei în cazul regresiilor multiple liniare. Metodele propuse, oricît de rafinate ar fi ele, încearcă să diminueze efectele multicoliniarităţii şi niciodată să le elimine [8].

În faza de stabilire a variabilelor modelului, dacă prin analiza calitativă s-a constatat că între doi sau mai mulţi factori există o relaţie obiectivă, atunci aceşti factori nu se reţin toţi ca factori de influenţă, ci se reduc, rămînînd factorii reciproc independenţi sau dependenţi în mică măsură. Problema, care anume factori să fie excluşi şi care să fie reţinuţi, se clarifică pe bază de considerente teoretice şi empirice. Formal, multicoliniaritatea se recunoaşte cu uşurinţă în valorile mari ale coeficienţilor de corelaţie simplă; de obicei, se ia drept limită valoarea 0,8 a coeficientului de corelaţie simplă pentru eliminarea unuia din cele două argumente coliniare[9]. Inexistenţa coliniarităţii se exprimă matematic sub forma egalităţii $\bar{x}_1\bar{x}_2 = \overline{x_1x_2}$, ceea ce nu se prea întîlneşte în practică. În orice caz, cu ajutorul exclusiv al analizei calitative nu se poate constata şi evita multicoliniaritatea. Este posibil

[7] Datorită suprapunerii influenţelor, coeficienţii de regresie apar deformaţi şi modelul în întregime îşi pierde valoarea. Dacă ne-am închipui legături riguros liniare între factori, consecinţa ar fi indeterminarea coeficienţilor de regresie.

[8] Problema multicoliniarităţii a fost abordată pentru prima dată de Ragnar Frisch în lucrarea *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934. El a propus o metodă de analiză a multicoliniarităţii, cunoscută sub denumirea de metoda fasciilor sau harta jetului radial *(bunch map)*. Esenţa metodei constă în calcularea coeficienţilor de regresie standardizaţi şi reprezentarea grafică a acestora pe o hartă cu jet radial. Prin examinarea proprietăţilor jetului radial se pot identifica relaţiile multicoliniare şi se poate aprecia dacă includerea uneia sau alteia dintre variabilele argument este utilă, inutilă sau dăunătoare. Este privită ca o metodă de determinare aproximativă a multicoliniarităţii, conţinînd impreciziile inerente caracterului său pur grafic.

[9] Un criteriu suplimentar ar fi mărimea coeficienţilor de corelaţie simplă, inferioară mărimii coeficientului de corelaţie multiplă, calculat în raport cu toate variabilele explicative. Fără îndoială, nu este vorba de criterii sigure, întrucît se au în vedere coeficienţii de corelaţie simplă, care, pe lîngă relaţiile între perechile respective de variabile explicative, exprimă implicit şi alte relaţii, contrar coeficienţilor de corelaţie parţială.

Scanned with OKEN Scanner

ca două argumente să fie independente din punctul de vedere al conținutului, dar datele numerice să exprime totuși o covariație și invers.

Din cauza intercorelațiilor caracteristice modelelor multifactoriale, cel puțin formal, dacă nu în fond, la care se adaugă dificultățile de interpretare, multe aplicații se rezumă la modele cu un singur factor. Construirea de modele monofactoriale este înlesnită, uneori, de posibilitatea reducerii numărului de factori prin adoptarea de indicatori statistici derivați sau înlocuirea unor indicatori cu alții. De exemplu, se elimină modificarea numărului populației prin indicatori calculați pe un locuitor, variația întinderii suprafeței agricole, arabile etc. prin indicatori pe unitate de suprafață, modificarea prețurilor prin înlocuirea indicatorilor în prețuri curente cu indicatori în prețuri comparabile ș.a. O corelație, să presupunem cvadruplă: consumul alimentar în funcție de venitul familiei, prețurile de vînzare și mărimea familiei poate fi redusă la o corelație pereche: consumul în prețuri comparabile pe o persoană și venitul pe o persoană. Limitarea domeniului de investigație la un singur factor este admisă cu condiția ca analiza calitativă să o sugereze, iar linia de regresie să o confirme, „explicînd" îndeajuns fenomenul studiat. Dacă nu se întîmplă așa, trebuie să se adauge neapărat și alte variabile explicative, pînă ce se va obține un model de corelație corespunzător. Astfel, analiza formării volumului și structurii cererii de mărfuri relevă ca factori de prim ordin: veniturile și prețurile. În profil static sau în profil dinamic, pentru perioadele în care prețurile prezintă variații neînsemnate, modelele se reduc la raportul cerere-venit.

d) Variabilele incluse în calculele de corelație comportă *specificarea și sub raportul formei de exprimare.* Caracterul de regulă complex al variabilelor economice dă posibilitatea exprimării aceleiași variabile într-o gamă mai mult sau mai puțin largă de indicatori. Este știut, de pildă, că nivelul productivității muncii industriale se poate măsura prin numeroși indicatori: potrivit cu producția — indicatori naturali, convențional-naturali, valorici, de muncă, iar potrivit cu cheltuielile de muncă — indicatori pe lucrător, muncitor, muncitor din activitatea de bază, unitate de timp lucrat (zi, oră). La rîndul lor, indicatorii valorici se calculează în chip diferit, pe baza producției globale, producției-marfă, producției nete, costului normativ al prelucrării etc. Avantajele și neajunsurile fiecărui indicator al productivității muncii au fost și sînt discutate amănunțit în publicațiile economice și statistice. Chiar și pentru o variabilă economică aparent simplă, cum ar fi cheltuielile materiale, considerate factor al venitului net realizat de cooperativele agricole de producție în una din aplicațiile noastre, nu există o singură modalitate de exprimare. Ele sînt susceptibile de exprimare în valoare absolută, ca pondere în producția globală sau ca mărime medie la 100 ha teren agricol.

Decizia în alegerea formelor de exprimare a variabilelor corelate necesită argumente de ordin teoretic. Principial, cel mai bun indicator este *indicatorul purificat de orice element în stare să estompeze influența reală a factorului acceptat.* Ca orientare, citățm modul de rezolvare a problemei în discuție, prilejuit de o aplicație concretă pe tema corelației între productivitatea muncii sociale, înzestrarea tehnică a muncii, înzestrarea electroenergetică a muncii și calificarea forței de muncă [125] [126]. Toate variabilele pot fi exprimate în mai multe feluri. Drept indicator al productivității muncii sociale s-a luat

volumul fizic al venitului național pe o persoană ocupată în sfera producției materiale, indicator admis în general în literatura de specialitate, avînd cea mai cuprinzătoare semnificație economică. Într-adevăr, în forma determinată, indicatorul productivității muncii sociale reflectă efectul economisirii nu numai a muncii vii, ci și a celei materiale. Înzestrarea tehnică a muncii s-a definit ca raport între fondurile fixe productive estimate în valoarea completă și numărul populației ocupate în ramurile producției materiale. S-a socotit mai corect să se ia valoarea completă, deoarece transferul valorii asupra producției nu este însoțit de pierderea, în aceeași proporție, a valorii de întrebuințare. În orice caz, includerea fondurilor fixe la valoarea rămasă ar fi afectat corelația reală între masa mijloacelor de muncă — în natură — și productivitatea muncii. Înzestrarea electroenergetică a muncii s-a exprimat în consumul de energie electrică pe o persoană ocupată în sfera producției materiale, indicator de asemenea acceptat din punct de vedere teoretic și practic pe considerentul că reflectă gradul de utilizare a puterii instalate. În ceea ce privește calificarea — factor greu de determinat statistic — , în lipsa unor indicatori cantitativi direcți s-au folosit, la scara economiei naționale, cheltuielile alocate de la bugetul de stat pentru învățămînt și pregătirea cadrelor. Neîndoielnic, cu toată argumentarea invocată, indicatorii adoptați sînt perfectibili. În consecință, problema alegerii formei de exprimare a variabilelor modelului de corelație trebuie rezolvată la nivelul realizărilor metodologice actuale.

e) Dacă am admis, fie și numai ipotetic, dependența unui fenomen de unul sau mai mulți factori, atunci apare întrebarea: ce model matematic să adoptăm ca să fie propriu dependenței cercetate și apt pentru analiză și, eventual, previziune? Din punct de vedere metodologic, problema se reduce la *stabilirea formei ecuației de regresie*, destinată să descrie relația dintre variabila dependentă și variabilele independente incluse în model. O soluție convenabilă presupune, în primul rînd, înțelegerea esenței și mecanismului legăturii, reprezentarea legității schimbării fenomenului modelat sub acțiunea factorilor aleși, iar după aceea, alegerea — din arsenalul funcțiilor matematice — a tipului de funcție adecvat legității sesizate. Din partea cercetătorului se pretinde deci studierea atentă a fenomenelor corelate și cunoașterea proprietăților diverselor funcții de modelare a corelațiilor.

Alături de cunoștințele teoretice, experiența practică și rezultatele cercetărilor analoage, la stabilirea ecuației de regresie, care să aproximeze cel mai bine forma reală a dependenței cercetate, concură o serie de procedee statistice. Cel mai simplu procedeu este reprezentarea grafică a distribuțiilor statistice corelate (corelograma).

Negreșit, judecățile despre tipul funcției de modelare, formulate în baza cercetării corelogramei, conțin imprecizii care țin de procedeul utilizat. Datorită acestui fapt, criteriul grafic nu trebuie luat drept unic criteriu de clarificare a formei legăturii, așa cum se întîmplă deseori în practica modelării corelațiilor. Forma legăturii sugerată de graficul statistic trebuie confirmată de criterii obiective (cum ar fi linia de regresie empirică și raportul F).

Este clar că, în alegerea funcției de modelare, se admit compromisuri cu modelul ideal din punct de vedere teoretic. De multe ori, numărul de variabile este limitat de posibilitatea obținerii datelor și de resursele disponibile pentru efectuarea calculelor. De asemenea, se recurge la funcții ușor acceptabile sub raport statistic, pentru estimare și verificare.

Scanned with OKEN Scanner

Desigur, eforturile reduse, cît și ușurința interpretării fac să se prefere funcțiile de modelare simple unor funcții complicate. În numele acestui deziderat, nu trebuie însă să se admită „convenții" de natură să altereze forma reală a dependentei cercetate.

f) În sfîrșit, este locul să amintim încă o dată două condiții cerute de orice calcul statistic: *omogenitatea datelor* și *un număr mare de observații*. Nu trebuie să se uite că în calculele de corelație se operează cu mărimi medii și că orice medie reflectă cu atît mai precis proprietățile tipice ale colectivității cu cît mai omogenă este colectivitatea. Caracterul omogen sau neomogen al colectivității, reprezentate prin datele statistice, poate fi sesizat cu ușurință pe calea examinării diagramei de dispersie a unităților de observare în raport cu valorile variabilelor corelate. Tendința punctelor de a se strînge în două sau în mai multe grupuri ori situarea unor puncte cu mult în afara norilor de puncte evidențiază eterogenitatea colectivității. În primul caz, soluția de adoptat ar fi dezmembrarea distribuțiilor statistice conform grupelor conturate, urmînd să se rezolve modelul pentru fiecare grupă obținută, iar în al doilea caz, eliminarea din calcule a observațiilor „aberante" față de masa observațiilor. Este vorba de „purificarea" corelației cercetate de influența altor factori, obiectiv prezenți în mediul fenomenului și de natură să modifice sensibil indicatorii de corelație. Sînt posibile modificări nu numai în gradul de corelație — corelație scăzută sau lipsa corelației în colectivitatea nediferențiată și corelație ridicată sau moderată în subcolectivitățile constituite și invers — dar, uneori, modificări chiar în ce privește sensul corelației, cum ar fi: corelație directă pe ansamblu și corelație inversă în una din subcolectivități [10].

O impresie înșelătoare despre intensitatea legăturii s-ar putea forma și în cazul menținerii valorilor cu abateri extreme în seriile de date statistice corelate. Prezența unei singure observații neomogene poate să schimbe mult parametrii ecuației de regresie și coeficientul (raportul) de corelație.

Mai trebuie avută în vedere și o altă situație. Cînd variația însăși a variabilelor corelate este întinsă, este indicată rezolvarea modelului de corelație atît la nivelul colectivității generale, cît și pe clase de unități, calitativ distincte. Este o modalitate de abordare comună tuturor indicatorilor statistici care sintetizează în cel mai înalt grad informațiile furnizate de o masă de cifre, cum sînt și indicatorii de corelație. Astfel, în scopul adîncirii cunoașterii proceselor economice, mediile sau indicii generali se completează cu medii, respectiv indici pe grupe.

Cît privește cealaltă condiție preliminară — un număr mare de observații — nesocotirea ei creează pericolul obținerii unor indicatori de corelație cu valori ridicate, dar nesiguri și nereprezentativi. În esență, acest lucru este bine cunoscut și nu facem decît să-l amintim.

În scopul asigurării autenticității estimațiilor, cîteodată se recurge la mărirea volumului de informații prin procedeul pe care l-am putea numi „unitate-ani": indicatorii fiecărei unități de observare se iau pentru o serie de ani, considerîndu-se rezultatul unor observații independente. O selecție, să zicem, de n unități observate într-o perioadă de m ani ar însemna un volum

[10] Se citează, ca instructiv pentru aplicațiile în orice domeniu, exemplul dat de K. Pearson cu privire la rezultatele cercetării corelației dintre lungimea și lățimea craniului: bărbați (806 observații) $r = 0{,}0869 \pm 0{,}0236$; femei (340 observații $r = -0{,}0424 \pm 0{,}0365$, iar pe cele două sexe considerate împreună: $r = 0{,}1968 \pm 0{,}0192$. Alte exemple, vezi [29].

de *nm* unităţi. Acest procedeu trebuie privit cu multă rezervă, deoarece poate duce la reflectarea deformată a corelaţiilor. Într-adevăr, indicatorii unui număr de unităţi-ani reprezintă, în fond, cu totul altceva decît indicatorii unui număr egal de unităţi pe un singur an. De fapt, datorită interdependenţei observaţiilor succesive, în primul caz avem de-a face numai cu *n* observaţii independente.

Atenţia trebuie îndreptată şi asupra unor condiţii de ordin tehnic, cum ar fi gruparea datelor statistice.

Prin operaţia de grupare se comprimă un mare volum de date, economisindu-se timpul de efectuare a calculelor de corelaţie, dar, în acelaşi timp, se sacrifică o anumită cantitate de informaţie. În consecinţă, valorile obţinute din date grupate diferă de valorile care s-ar fi obţinut dacă calculele de corelaţie s-ar fi efectuat pe baza datelor negrupate.

În legătură cu gruparea este discutabilă problema aplicării calculelor de corelaţie la repartiţii statistice cu intervale neegale. Regula generală, motivată teoretic şi confirmată de experienţa rezolvării modelelor de corelaţie, constă în folosirea repartiţiilor cu intervale egale. Cum însă această regulă este frecvent încălcată în materialele statistice gata elaborate, soluţia ar putea fi rezolvarea modelului de corelaţie pe diviziuni cu intervale egale. În caz contrar, trebuie să se recurgă la gruparea datelor primare conform regulii ştiute, pentru a garanta — cu satisfacerea şi a tuturor celorlalte condiţii — estimaţii ale indicatorilor de corelaţie suficient de sigure.

8.4 REGRESIA

După cum s-a precizat, avem de-a face cu o legătură de tip statistic, cînd, la variaţia uneia din variabile, cealaltă variabilă reacţionează prin schimbarea repartiţiei. Vizual, această variaţie se poate constata pe tabelul de corelaţie (tabelul 8.2) şi, mai expresiv, pe diagrama de dispersie (fig. 8.1).

Dacă pentru fiecare repartiţie y calculăm valoarea ei medie, rezultă o serie de asemenea valori care, ştim, sînt medii condiţionate ($\hat{y}_x$) de valorile pe care le ia variabila x (tabelul 8.3). Reprezentarea grafică pe un sistem de axe rectangulare a perechilor de valori x, $\hat{y}_x$ şi unirea punctelor obţinute descrie o linie poligonală numită *linie de regresie* [11] a variabilei y în raport cu variabila x (fig. 8.1.).

Se înţelege că analog determinării liniei de regresie a lui y în x se poate determina linia de regresie a lui x în y (tabelul 8.3 şi fig. 8.1.). În cazul în care

[11] Termenul „regresie" (lat. *regressio* = întoarcere) a fost introdus în statistică de F. Galton. Cercetînd problemele eredităţii pe baza observaţiilor statistice, el a descoperit, printre altele, că din părinţi a căror talie este mai mică decît media colectivităţii provin copii cu o talie superioară lor şi invers. Iată „mediile legate" rezultate din prelucrarea observaţiilor asupra a 928 cazuri, conform tabelului cu dublă intrare construit de Galton:

	sub									*peste*
x	64,5	64,5	65,5	66,5	67,5	68,5	69,5	70,5	71,5	72,5	72,5
$\hat{y}$	65,0	65,3	66,7	67,1	68,1	68,0	68,7	68,2	68,1	67,2	72,9,

în care x reprezintă talia părinţilor (ţoli), $\hat{y}$ — talia medie a copiilor, corespunzătoare taliei

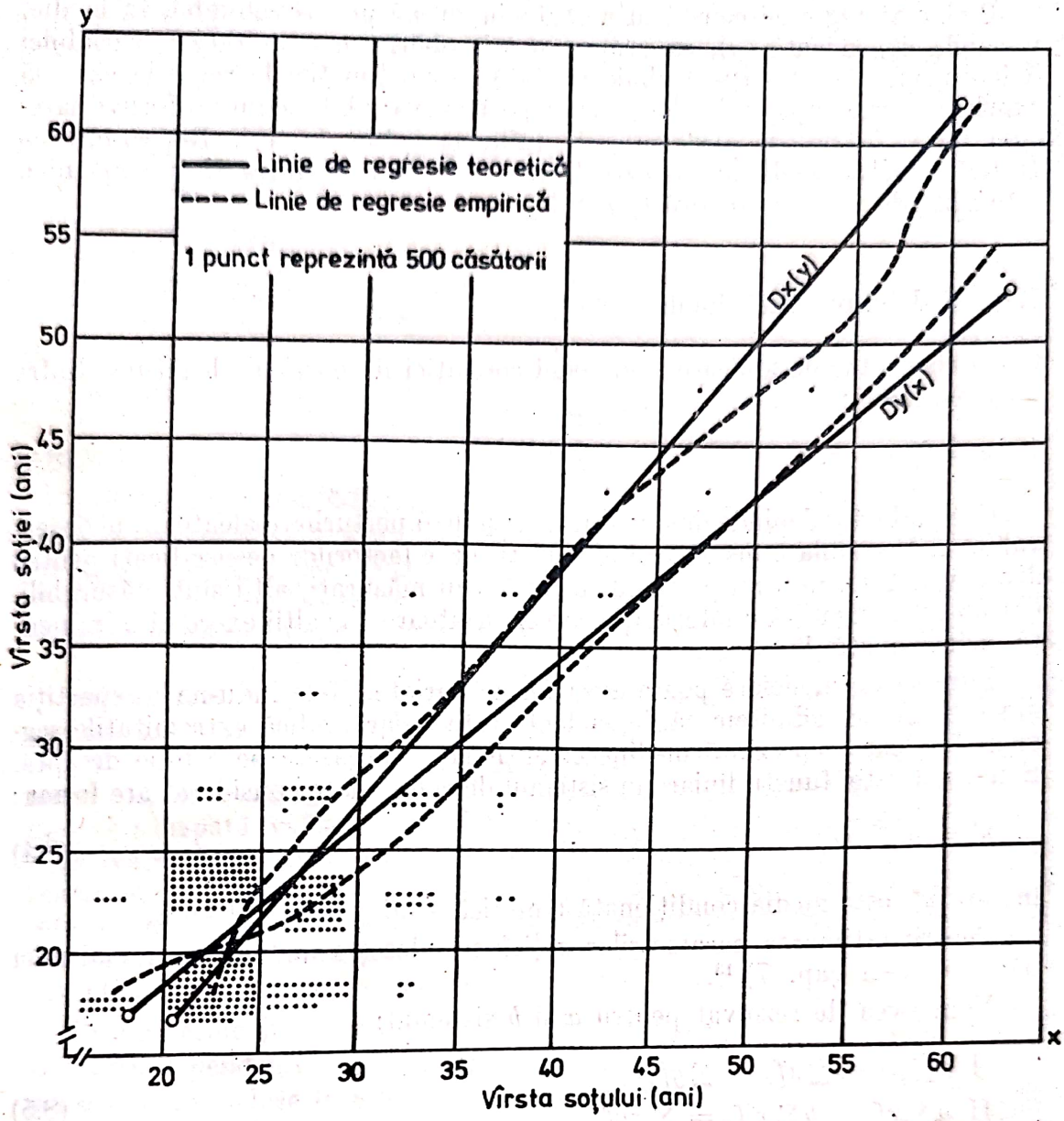


Fig. 8.1 Corelația dintre vîrsta soțului și vîrsta soției în căsătoriile din R. S. România, în anul 1974.

variabilele x și y sînt independente, linia de regresie a variabilei y în x este o dreaptă paralelă cu axa absciselor, iar linia de regresie a variabilei x în y — o dreaptă paralelă cu axa ordonatelor.

părinților din fiecare grupă (țoli). Talia medie a părinților în colectivitatea celor 982 observații este egală cu 68,3 țoli.

Prin urmare, se constată o regresie către valoarea medie, de unde și denumirea linie de regresie dată de Galton liniei descrise de reprezentarea grafică a perechilor de valori $x_i \hat{y}_i$ într-un sistem de axe rectangulare.

Termenul regresie fiind impropriu pentru cazul general, s-a propus un termen mai potrivit, și anume „estimare" *(estimating ecuation)*, denumire folosită încă în anul 1909 de G. Udny Yule (vezi [133]). Totuși denumirile: analiză de regresie, metodă de regresie, ecuație de regresie, linie de regresie, coeficient de regresie sînt uzuale.

Problema regresiei constă în a arăta în ce măsură se schimbă, în medie, variabila dependentă (y), ca urmare a schimbării cu o unitate a variabilei independente (x). Soluția trebuie căutată într-o funcție de regresie care să exprime dependența variabilei y în raport cu variabila x într-o formă oarecum pură, făcîndu-se abstracție de influența altor factori. Reprezentarea grafică a unei astfel de funcții, care descrie matematic legătura dintre variabile, este *linia de regresie teoretică* (fig. 8.1.)

8.4.1 Modelul regresiei simple

Cazul elementar de aplicare a metodei corelației îl constituie legătura dintre două variabile:

$$y = f(x) + \varepsilon. \tag{8.3}$$

O variabilă y depinde de o variabilă x și de o perturbare aleatoare neobservabilă ε. Variabila ε însumează efectul tuturor factorilor nespecificați pentru diverse motive: unii nu pot fi identificați sau măsurați, alții sînt măsurabili, dar nu este asigurată o informație corespunzătoare, iar alții exercită o influență neînsemnată. [12]

Funcția de modelare poate avea forme variate. Referindu-ne la repartiția din tabelul 8.3, admitem că dependența este liniară, adică extremitățile segmentelor, care reprezintă mediile condiționate, se găsesc pe o linie dreaptă. Ecuația acestei funcții liniare în sistemul de coordonate considerat are forma:

$$y_x^* = a + bx, \tag{8.4}$$

în care y^* este media condiționată teoretică a lui y în raport cu x.

Pentru estimarea parametrilor a și b se folosește metoda celor mai mici pătrate (vezi cap. 7) [13].

Vom avea de rezolvat pentru a și b sistemul:

$$\begin{aligned} &\text{I} \quad a\Sigma f + b\Sigma x f_x = \Sigma y f_y \\ &\text{II} \quad a\Sigma x f_x + b\Sigma x^2 f_x = \Sigma xy f_{xy} \end{aligned} \tag{8.5}$$

Introducînd în sistem sumele calculate pe baza tabelului 8.2 obținem:

$$\text{I} \quad 175496\ a + 4571582\ b = 3947137$$

$$\text{II} \quad 4571582\ a + 130044744\ b = 111413429.$$

[12] Termenul este considerat aleator, deoarece ignoranța noastră asupra modalităților sale de determinare îi conferă natura unei variabile aleatoare. Ca urmare a relației (8.3), mărimea y este aleatoare, ca și variabila ε.

[13] Metoda celor mai mici pătrate, ca metodă clasică în rezolvarea modelelor de corelație răspunde mai bine ideii de a se aduce la medie influența factorilor specificați în model, precum și influența de ansamblu a factorilor neluați în considerare. Parametrii estimați sînt mărimi medii și tot medii sînt și valorile estimate ale variabilei dependente. În întregime, ecuația de regresie exprimă o relație univocă, însă aceasta este o relație stohastică adusă la medie. Metoda celor mai mici pătrate prezintă și avantajul că, pentru o largă clasă de funcții, nu comportă scheme de calcul prea complicate.

Scanned with OKEN Scanner

Soluția sistemului cu ajutorul determinanților:

$$a = \frac{\begin{vmatrix} \Sigma y f_y & \Sigma x f_x \\ \Sigma xy f_{xy} & \Sigma x^2 f_x \end{vmatrix}}{\begin{vmatrix} \Sigma f & \Sigma x f_x \\ \Sigma x f_x & \Sigma x^2 f_x \end{vmatrix}} = \frac{\Sigma y f_y \Sigma x^2 f_x - \Sigma xy f_{xy} \Sigma x f_x}{\Sigma f \Sigma x^2 f_x - \Sigma x f_x \Sigma x f_x}; \tag{8.6}$$

$$b = \frac{\begin{vmatrix} \Sigma f & \Sigma y f_y \\ \Sigma x f_x & \Sigma xy f_{xy} \end{vmatrix}}{\begin{vmatrix} \Sigma f & \Sigma x f_x \\ \Sigma x f_x & \Sigma x^2 f_x \end{vmatrix}} = \frac{\Sigma f \Sigma xy f_{xy} - \Sigma x f_x \Sigma y f_y}{\Sigma f \Sigma x^2 f_x - \Sigma x f_x \Sigma x f_x}; \tag{8.7}$$

$$a = \frac{(3947137 \times 130044744) - (111413429 \times 4571582)}{(175496 \times 130044744) - (4571582)^2} = 2{,}0659;$$

$$b = \frac{(175496 \times 111413429) - (4571582 \times 3947137)}{(175496 \times 130044744) - (4571582)^2} = 0{,}7841.$$

Ecuația regresiei [14]

$$y^* = 2{,}0659 + 0{,}7841x.$$

Parametrul b, numit coeficient de regresie (sau coeficient unghiular al regresiei), reprezintă, în sens geometric, panta liniei drepte. În calculul corelației, el arată cu cît se modifică variabila y, cînd x se modifică cu o unitate, adică răspunde tocmai la problema regresiei. În cazul corelației directe, parametrul b capătă valoare pozitivă, iar în cazul corelației inverse — valoare negativă [15]. Parametrul liber a nu are, de regulă, nici un fel de semnificație independentă, ci una de calcul, putînd avea sens negativ. Uneori, se încearcă interpretarea sensului acestui parametru ca valoarea mărimii y pentru $x = 0$. O asemenea interpretare duce, în multe cazuri, la absurdități, fiind imposibil să-i atribuim o valoare materială oarecare.

Prin intermediul parametrilor ecuației de regresie se estimează valorile variabilei dependente pentru valorile cunoscute ale variabilei sau variabilelor independente incluse în modelul de corelație. Dacă variabilele și forma legăturii s-au ales adecvat și dacă datele statistice utilizate la calculul ecuației satisfac condiția de omogenitate și condiția de volum, atunci este de așteptat ca relația ajustată $y = f(x)$ să reprezinte cu maximum de verosimilitate relația $y = \hat{y}$.

Exemplul 2. Să examinăm estimațiile vînzărilor de mărfuri cu amănuntul prin comerțul socialist (y) în funcție de veniturile bănești ale populației (x), pe baza unui model dinamic.

[14] Ecuația de regresie, considerînd vîrsta soțului (x) în funcție de vîrsta soției (y) este: $x^* = 6{,}6687 + 0{,}8617y$

(elementele de calcul în tabelul 8.2).

[15] Cînd variațiile a două fenomene se produc în același sens, se spune că corelația este *directă* sau *pozitivă*, iar cînd variațiile sînt de sens contrar, corelația este numită *inversă* sau *negativă*.

Scanned with OKEN Scanner

Tabelul 8.4

Anii	Venituri băneşti (%) x	Vînzări de mărfuri (mlrd. lei) y	y^*	$y - y^*$	$(y - y^*)^2$
1960	100,0	41,5	43,5	−2,0	4,00
1961	111,7	47,2	47,9	−0,7	0,49
1962	125,3	53,0	53,0	0	0
1963	137,2	57,5	57,4	+0,1	0,01
1964	147,4	62,8	61,3	+1,5	2,25
1965	163,2	67,6	67,2	+0,4	0,16
1966	178,8	74,3	73,0	+1,3	1,69
1967	195,3	81,4	79,2	+2,2	4,84
1968	212,1	84,2	85,5	−1,3	1,69
1969	223,7	88,7	89,8	−1,1	1,21
1970	242,6	96,7	96,9	−0,2	0,04
1971	263,8	105,9	104,9	+1,0	1,00
1972	284,2	112,4	112,5	−0,1	0,01
1973	310,6	121,3	122,4	−1,1	1,21
Total	—	1094,5	1094,5	0	18,60

Sursa: *Anuarul statistic al Republicii Socialiste România 1970, 1972, 1974.*

Ecuaţia de regresie:

$$y_x^* = 6{,}0241173 + 0{,}3747032x. \tag{8.8}$$

Dacă potrivit relaţiei [16]

$$b_{abs.} = b_{rel.} \frac{100}{x} \tag{8.9}$$

se transformă în valoare absolută coeficientul de regresie, se obţine 0,6952.

O creştere a veniturilor băneşti, în perioada 1960—1973, în medie cu 100 lei a generat creşterea cererii de mărfuri cu circa 70 lei.

Vînzările de mărfuri estimate cu ajutorul ecuaţiei de regresie se abat de la vînzările efective, ceea ce constituie consecinţa firească a caracterului statistic al legăturii. Abaterile înregistrate sînt efectul acţiunii altor factori ai cererii, de care s-a făcut abstracţie.

Calculînd regresia teoretică, am obţinut, ca formă, schimbarea univocă a volumului vînzărilor de mărfuri odată cu schimbarea veniturilor băneşti. Ecuaţia de regresie exprimă aspectul statistic al legăturii şi, ca atare, ea are eficacitate pentru ansamblul observaţiilor, şi nu pentru fiecare observaţie în parte. În reprezentarea grafică, linia de regresie poate să urmeze doar direcţia desfăşurării punctelor, fără însă să coincidă cu acestea. Este absolut evident că nu se pot alege valorile parametrilor (şi nici nu se pune aşa problema), astfel încît curba să treacă prin toate punctele (fig. 8.2).

Modelul liniar (8.4) semnifică o schimbare uniformă a variabilei dependente sub influenţa schimbării variabilei independente [17]. Oricare ar fi nivelul varia-

[16] În seriile cronologice corelate, veniturile băneşti au fost exprimate în indici cu bază constantă (1960 = 100), iar vînzările de mărfuri, în valori absolute. În consecinţă, parametrul a s-a obţinut în valoare absolută, iar b, în valoare relativă.

[17] Pentru funcţia liniară (8.4), proporţionalitatea directă există între $y - a$ şi x, şi nu între y şi x.

Scanned with OKEN Scanner

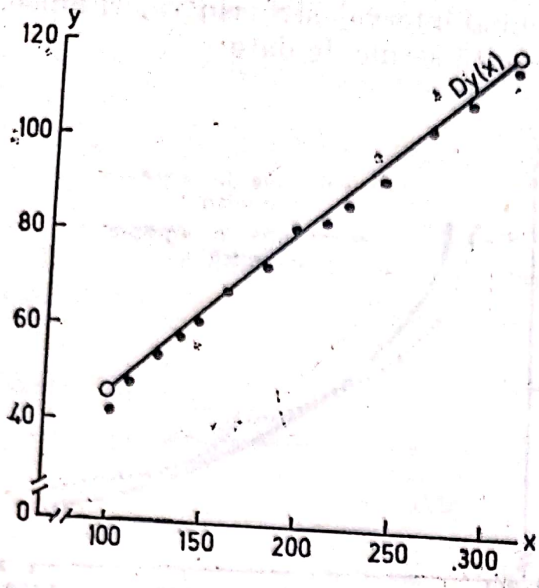


Fig. 8.2 Corelația dintre volumul vînzărilor de mărfuri și veniturile bănești ale populației.

bilei independente, creșterea sau descreșterea ei cu o unitate se consideră că produce, în fiecare caz, creșterea sau descreșterea variabilei dependente cu aceeași mărime (mărimea b — coeficientul de regresie). Funcția liniară își găsește aplicația și cînd variațiile variabilelor nu urmează legea progresiei aritmetice, dar ele sînt mici și deci curba descrisă este foarte aproape de linia dreaptă.

În cercetările practice se întîlnesc destul de des legături statistice, în care intensitatea influenței variabilei independente asupra variabilei dependente depinde de însuși nivelul la care se află variabila independentă.

Exemplul 3. Să ne referim, de exemplu, la corelația dintre prețul de cost (y) și producția medie la hectar de grîu și secară (x), conform datelor din tabelul 8.5.

Tabelul 8.5

y \ x	1400—1600	1600—1800	1800—2000	2000—2200	2200—2400	2400—2600	2600—2800	2800—3000	3000—3200	3200—3400	3400—3600	3600—3800	f_y
1400—1500	1												1
1300—1400	1	1											2
1200—1300	2	2											4
1100—1200		6	1	1	2								10
1000—1100		3	12	5	2	4							26
900—1000		8	11	11	7	4	4	2					47
800—900		1	2	10	15	10	7	1	1				47
700—800			1	6	12	15	13	8	2	1	1		59
600—700					1	7	8	10	4	4	1	1	36
500—600						1		3	3	2	2	1	12
f_x	4	21	27	33	39	41	32	24	10	7	4	2	244

Notă: Sînt grupate 244 întreprinderi agricole de stat, care au cultivat grîu și secară în anii 1967—1969. S-au calculat valori medii pentru cei trei ani: y — cheltuieli medii de producție (lei/tonă) și x — producția medie (kg/hectar).

Calculînd mediile legate ale lui y și înlocuind intervalele x prin valori unice, situate la mijlocul scării lor de variație, rezultă seriile de date:

x	$\hat{y}_x$	x	$\hat{y}_x$
1500	1325	2700	772
1700	1064	2900	704
1900	987	3100	660
2100	905	3300	636
2300	858	3500	625
2500	801	3700	600

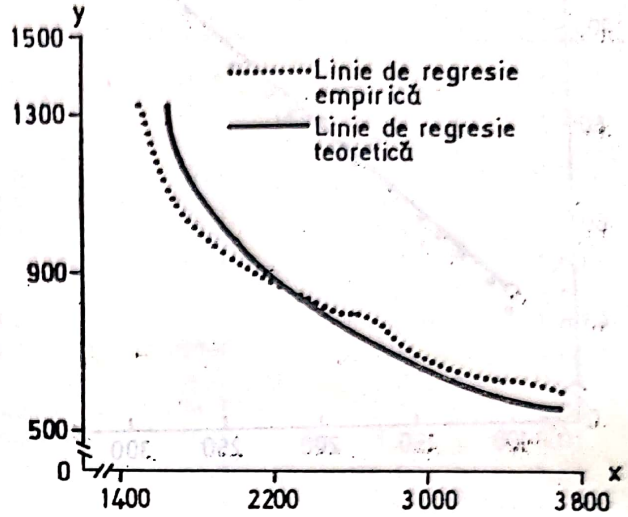


Fig. 8.3

Reprezentarea grafică a perechilor de valori x_i, $\hat{y}_i$ înscrie o linie poligonală, care evidențiază deosebit de limpede *caracterul neliniar* al legăturii dintre cele două variabile (fig. 8.3.).

Ipoteza unei *legături de tip hiperbolic*

$$y^* = a + b\frac{1}{x} \tag{8.10}$$

cu care se pornește — potrivit teoriei economice — la descrierea relației stohastice preț de cost/producție se verifică de raportul definit de formula [18]

$$F_{calc} = \frac{\Sigma(\hat{y} - y^*)^2 f_{xy}}{k-2} : \frac{\Sigma(y - \hat{y})^2 f_{xy}}{\Sigma f - k}. \tag{8.11}$$

Tabelul 8.6

	Variația care corespunde hiperbolei*	Variația care nu corespunde hiperbolei*	Variația în interiorul grupelor
Suma pătratelor abaterilor	4195191,3	91251,2	4285508,2
Numărul gradelor de libertate		10	232
Dispersia		9125,12	11777,88
F_{calc}			0,77

* $y^* = 179{,}826 + \frac{1544{,}885}{x}$. (8.12)

[18] Formula (8.11) exprimă raportul dintre dispersia față de linia de regresie adoptată și dispersia în interiorul grupelor. Raportul F urmează o repartiție Behrens — Fisher — Snedecor cu $k-2$ și $\Sigma f - k$ grade de libertate (k reprezintă numărul grupelor după variabila x).

Scanned with OKEN Scanner

$F_{calc.} < F_{0,99} \ldots 0,77 < 2,41.$

S-a probat și ipoteza legăturii liniare, care însă nu se verifică [19] deoarece:

$F_{calc.} > F_{0,99} \ldots 2,82 > 2,41.$

În favoarea adoptării funcției de modelare hiperbolică, testată de raportul F, intervine și eroarea standard a evaluării, după cum reiese din tabelul 8.7.

Tabelul 8.7

Corelația	Suma pătratelor abaterilor de la		Numărul i.a.s.	Eroarea standard a evaluării			
	hiperbolă	dreaptă		hiperbolă		dreaptă	
				lei	%	lei	%
Prețul de cost — producția medie de grîu și secară	91251,2	332677,1	244	19,34	2,3	36,92	4,4

Notă. Eroarea standard a evaluării s-a calculat după formula:

$$s_{xy} = \sqrt{\frac{\Sigma (\hat{y} - y^*)^2 f_{xy}}{\Sigma f}} \quad (8.13)$$ în valoare absolută și în valoare relativă

$$\frac{s_{yx}}{\bar{y}} \cdot 100. \quad (8.14)$$

Despre relația dintre variațiile producției și prețului de cost, luată ca exemplu, se poate gîndi, de la bun început, în felul următor: odată cu creșterea volumului producției — celelalte condiții păstrîndu-se egale — prețul de cost scade, dar tendința de scădere nu are caracter uniform, datorită tendințelor diferite în modificarea celor două componente ale cheltuielilor de producție: cheltuielile fixe și cheltuielile variabile. Totdeauna prețul de cost se reduce mai încet decît crește producția și acest decalaj se amplifică cu cît cheltuielile fixe pe unitatea producătoare dețin o pondere mai mică în ansamblul cheltuielilor de producție.

Examinarea cîmpului de dispersie ca și a liniei de regresie empirică (fig. 8.3.) ne arată o tendință de scădere a prețului de cost mai accentuată la valorile inițiale ale repartiției, devenind, după o anumită sporire a volumului producției, mai lentă.

Ecuația (8.12) arată că la 1 tonă producție de grîu-secară la hectar reveneau, în anii considerați, aproximativ 180 lei cheltuieli variabile și în

[19] Ecuația de regresie liniară:
$y^* = 1511,53 - 0,2764\ x.$

plus o anumită parte din suma de 1545 lei, cheltuieli convenţional-constante. Cheltuielile medii de producţie pe o tonă erau:

— la producţia medie minimă (1,5 tone)

$$179{,}826 + \frac{1544{,}885}{1{,}5} = 1209{,}7 \text{ lei};$$

— la producţia medie maximă (3,7 tone)

$$179{,}826 + \frac{1544{,}885}{3{,}7} = 597{,}4 \text{ lei}.$$

Dacă ecuaţia (8.12) este luată abstract, cu considerarea unei creşteri nelimitate a producţiei medii la hectar, atunci partea ce ar reveni pe tona de producţie din cheltuielile convenţional-constante ar tinde către o valoare infimă şi cheltuielile de producţie pe tonă s-ar apropia foarte mult de 180 lei.

Multe legături statistice se înscriu pe curbe definite de ecuaţiile de regresie:

$$y^* = a + bx + b'x^2 \tag{8.15}$$

$$y^* = ax^b. \tag{8.16}$$

Parabola (8.15) este specifică exprimării dependenţelor cu o alură ce tinde către un punct maxim (minim).

Exemplul 4. Corelaţia dintre producţia de porumb (y) şi densitatea plantelor (x):

Tabelul 8.8

Numărul de plante la hectar (mii)	Producţia de boabe la hectar (chintale) obţinută pe lotul:				
	1	2	3	4	5
20	41,8	40,5	43,9	42,6	44,7
30	49,3	47,5	48,1	45,8	51,8
40	52,3	52,7	51,5	49,1	51,0
50	50,1	56,7	53,7	58,1	54,9
60	55,0	53,8	57,4	50,3	54,0
70	53,3	53,1	51,6	51,2	49,3

Sursa: *Experienţe cu porumb dublu hibrid*, Editura Agrosilvică, Bucureşti, 1962, p. 107. (Repartiţiile pe loturi sînt imaginate de noi.)

Teoretic şi logic, corelaţia dintre recolta de porumb şi o serie de factori de influenţă, cum sînt: densitatea plantelor, cantitatea de îngrăşăminte, regimul de irigare etc., îmbracă aproximativ forma unei parabole. În speţă, experienţa acumulată prin analiza datelor statistice de masă demonstrează că, pe măsura creşterii densităţii plantelor, creşte şi recolta, dar numai pînă la o anumită limită. După acest punct critic, densitatea plantelor devine un factor negativ şi recolta scade.

Scanned with OKEN Scanner

Sistemul de ecuaţii normale pentru parabola de gradul doi este:

$$\begin{aligned} &\text{I}\quad na + b\Sigma x + b'\Sigma x^2 = \Sigma y \\ &\text{II}\quad a\Sigma x + b\Sigma x^2 + b'\Sigma x^3 = \Sigma xy \\ &\text{III}\quad a\Sigma x^2 + b\Sigma x^3 + b'\Sigma x^4 = \Sigma x^2 y. \end{aligned} \tag{8.17}$$

Substituind valorile numerice ale sumelor corespunzătoare [20], rezultă:

$$\begin{aligned} &\text{I}\quad 6\ a + 270\ b + 13900\ b' = 303 \\ &\text{II}\quad 270\ a + 13900\ b + 783000\ b' = 13961 \\ &\text{III}\quad 13900\ a + 783000\ b + 46750000\ b' = 727650. \end{aligned} \tag{8.17'}$$

Ca metodă de rezolvare, vom folosi metoda simplex.

Operaţiile de calcul:

a) scrierea coeficienţilor sistemului de ecuaţii normale sub formă de tabel:

Tabelul 8.9

	a	b	b'	ω
a	[6]	270	13 900	303
b	270	13 900	783 000	13 961
b'	13 900	783 000	46 750 000	727 650
ω	303	13 961	727 650	15 399,02*

* Σy^2.

b) eliminarea coloanei a: fiecare coeficient al primei ecuaţii a fost împărţit la coeficientul a al acesteia, apoi din coeficienţii ecuaţiilor următoare s-a scăzut mărimea rezultată din împărţirea coeficientului corespunzător al primei ecuaţii la coeficientul a şi înmulţită cu primul coeficient al ecuaţiei respective (tabelul 8.10).

Tabelul 8.10

	a	b	b'	ω
a	1	45	2 316,66	50,5
b	0	[1 750]	157 500	326
b'	0	157 500	14 548 333,43	27 500
ω	0	326	25 700	97,52*

* $\Sigma(y-\overline{y})^2$.

[20] Deoarece numărul de observaţii în fiecare grupă este acelaşi (tabelul 8.8), s-a operat cu seriile de valori:

x	20	30	40	50	60	70
y	42,7	48,5	51,3	54,7	54,1	51,7.

Scanned with OKEN Scanner

c) eliminarea coloanei b: elementele rîndului al doilea al tabelului 8.10 s-au împărţit cu primul element al acestui rînd după care elementele celorlalte rînduri s-au scăzut cu cîtul corespunzător înmulţit cu primul element al rîndului respectiv.

Tabelul 8.11

	b	b'	ω
a	0	−1733,33	42,117
b	1	90	0,186
b'	0	373333,43	−3460
ω	0	−3640	36,79

d) elementele rîndului al treilea al tabelului 8.11 s-au împărţit cu primul element al său, apoi procedîndu-se în mod analog cu cele de la pct. b şi c, au rezultat valorile parametrilor a, b şi b' (tabelul 8.12).

Tabelul 8.12

a	25,21757
b	1,06377
b'	−0,00975
ω	1,33640*

* $\Sigma(y - y^*)^2$.

Datele experimentale ne conduc la ecuaţia de regresie.

$$y^* = 25{,}217 + 1{,}0637\ x - 0{,}00975\ x^2, \qquad (8.18)$$

faţă de care dispersia valorilor observate este cu mult mai redusă decît în cazul ecuaţiei liniei drepte (tabelul 8.13)

Tabelul 8.13

	Variaţia care corespunde		Variaţia care nu corespunde		Variaţia în interiorul grupelor
	parabolei	dreptei*'	parabolei	dreptei*	
Suma pătratelor abaterilor	481,06	303,96	6,45	183,64	113,55
Numărul gradelor de libertate			3	4	24
Dispersia			2,18	45,91	4,73

* $y^* = 42{,}117 + 0{,}1863\ x$.

Raportul F calculat pentru:

— parabolă $\frac{2{,}18}{4{,}73} = 0{,}46$;

— dreaptă $\frac{45{,}91}{4{,}73} = 9{,}71$,

deoarece $F_{calc.} = 0{,}46 < F_{0,99} = 4{,}72$, ipoteza legăturii parabolice se confirmă, iar $F_{calc.} = 9{,}71 > F_{0,99} = 4{,}22$, ipoteza legăturii liniare nu se confirmă.

Scanned with OKEN Scanner

Funcţia (8.18) prezintă un mare interes practic, întrucît cu ajutorul ei se determină uşor „corelaţia optimă" între cele două variabile. În intervalul de variaţie al datelor empirice, funcţia atinge un punct maxim, definit de $x_{max} = 55$ mii plante la hectar şi $y_{max} = 54$ chintale la hectar [21]. Dincolo de „densitatea maximă" — celelalte condiţii presupuse egale — recolta tinde să scadă.

Funcţia putere (8.16) implică un ritm de variaţie constant, de unde şi valabilitatea ei numai pentru fenomene cu evoluţie continuă, neadmiţînd — cel puţin în cuprinsul perioadei de analiză sau şi de previziune — nivel de saturaţie. Aproximarea exponenţială, ca şi cea liniară, apare potrivită în cazul intervalelor de variaţie nu prea mari; ea este recomandată şi atunci cînd nu se cunoaşte forma dependenţei într-un mod mai precis. Avantajul elasticităţii constante şi al simplităţii calculului explică generalizarea aplicării curbelor cererii de tipul (8.16).

Exemplul 5. Să rezolvăm modelul (8.16) pentru vînzările de mărfuri (y) în funcţie de veniturile băneşti ale populaţiei (x). În tabelul 8.14 se dau logaritmii valorilor x şi ai valorilor y din tabelul 8.4.

Tabelul 8.14

Anii	log x	log y	log y^*	y^* (mlrd. lei)
1960	2,00000	1,61805	1,63057	42,7
1961	2,04805	1,67394	1,67518	47,3
1962	2,09795	1,72428	1,72151	52,7
1963	2,13735	1,75967	1,75809	57,3
1964	2,16850	1,79796	1,78701	61,2
1965	2,21272	1,82995	1,82807	67,3
1966	2,25237	1,87099	1,86488	73,3
1967	2,29070	1,91062	1,90047	79,6
1968	2,32654	1,92531	1,93374	85,9
1969	2,34967	1,94792	1,95522	90,2
1970	2,38489	1,98543	1,98792	97,3
1971	2,42127	2,02490	2,02169	105,1
1972	2,45362	2,05077	2,05173	112,6
1973	2,49220	2,08386	2,08755	122,3
Total		26,20365	26,20363	1094,8

[21] Nivelul variabilei x la care variabila y capătă valoarea maximă se determină cu ajutorul primei derivate $b + 2\,b'x$ ajustată la zero. Cu datele funcţiei (8.18):

$$1{,}0637 + 2\ (-0{,}00975)\ x = 0,$$

de unde

$$x^* = \frac{1{,}0637}{2 \times 0{,}00975} = 54{,}5 \text{ mii plante}$$

şi

$$y^* = 25{,}27 + (1{,}0637 \times 54{,}5) - (0{,}00975 \times 54{,}5^2) = 54{,}28 \text{ chintale.}$$

Scanned with OKEN Scanner

Prin logaritmare, funcţia (8.16) devine liniară

$$\log y^* = \log a + b \log x \tag{8.19}$$

şi constantele $\log a$ şi b se obţin prin rezolvarea sistemului de ecuaţii normale:

$$\text{I } n \log a + b\sum \log x = \sum \log y \tag{8.20}$$

$$\text{II } \log a\sum \log x + b\sum (\log x)^2 = \sum (\log x \log y).$$

Cu datele cunoscute:

$$\text{I } 14 \log a + 31{,}6358300\ b = 26{,}2036500$$

$$\text{II } 31{,}6358300 \log a + 71{,}7984649\ b = 59{,}5011061.$$

Rezolvarea simultană dă:

$$\log a = -0{,}2263002;\ b = 0{,}9284363.$$

Ecuaţia de regresie

$$\log y^* = -0{,}2263002 + 0{,}9284363 \log x,$$

exprimată sub forma iniţială: [22]

$$y^* = 0{,}5939^{0{,}9284}. \tag{8.21}$$

În calitate de coeficient mediu de elasticitate, parametrul b calculat arată că, în perioada 1960—1973, la o creştere a veniturilor cu 10%, cererea a crescut, în medie, cu 9,3%.

8.4.2 Modelul regresiei multiple

Principiile regresiei simple se aplică şi în cazul regresiei multiple, dar calculele sînt mai laborioase. Modelul are următoarea formulare generală:

$$y = f(x_1, x_2, \ldots, x_n) + \varepsilon. \tag{8.22}$$

O variabilă endogenă y depinde de o serie de variabile exogene explicite $x_1, x_2 \ldots, x_n$ şi de o perturbare aleatoare neobservabilă ε.

Funcţiile de mai multe variabile, cel mai des utilizate, sînt de forma:

$$y^* = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n \tag{8.23}$$

sau

$$y^* = ax_1^{b_1}x_2^{b_2} \ldots x_n^{b_n}. \tag{8.24}$$

Pentru ecuaţia (8.23), aplicarea metodei celor mai mici pătrate echivalează cu condiţia minimizării expresiei:

$$\min \sum \varepsilon^2 = \sum (y - y^*)^2 = \sum [y - (a + b_1x_1 + b_2x_2 + \ldots + b_nx_n)]^2 \tag{8.25}$$

[22] De observat că $\log a = -0{,}2263002 = 0{,}7736998 - 1$, iar antilogaritmul său este 0,5939.

Scanned with OKEN Scanner

Anulînd derivatele parțiale ale expresiei (8.25) în raport cu parametrii $a, b_1, b_2, \ldots, b_n$

$$\frac{\partial}{\partial a}\varepsilon^2 = 2\Sigma(y - a + b_1x_1 + b_2x_2 + \ldots + b_nx_n)\cdot 1 = 0$$

$$\frac{\partial}{\partial b_1}\varepsilon^2 = 2\Sigma(y - a + b_1x_1 + b_2x_2 + \ldots + b_nx_n)\cdot x_1 = 0$$

$$\frac{\partial}{\partial b_2}\varepsilon^2 = 2\Sigma(y - a + b_1x_1 + b_2x_2 + \ldots + b_nx_n)\cdot x_2 = 0 \qquad (8.26)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{\partial}{\partial b_n}\varepsilon^2 = 2\Sigma(y - a + b_1x_1 + b_2x_2 + \ldots + b_nx_n)\cdot x_n = 0$$

rezultă sistemul de ecuații normale

$$na + b_1\Sigma x_1 + b_2\Sigma x_2 + \ldots + b_n\Sigma x_n = \Sigma y$$

$$a\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 + \ldots + b_n\Sigma x_1x_n = \Sigma yx_1$$

$$a\Sigma x_2 + b_1\Sigma x_1x_2 + b_2\Sigma x_2^2 + \ldots + b_n\Sigma x_2x_n = \Sigma yx_2 \qquad (8.27)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a\Sigma x_n + b_1\Sigma x_1x_n + b_2\Sigma x_2x_n + \ldots + b_n\Sigma x_n^2 = \Sigma yx_n.$$

Rezolvarea sistemului (8.27) cu datele cunoscute ne dă valorile căutate ale parametrilor ecuației de regresie (8.23).

Exemplul 6. Ne vom referi la o aplicație în care intervin trei variabile explicative: cercetarea dependenței venitului net realizat de cooperativele agricole de producție (y mii lei la 100 ha teren agricol) de factorii: [23]

x_1 — cheltuielile materiale (mii lei la 100 ha teren agricol);

x_2 — valoarea mijloacelor fixe (mii lei la 100 ha teren agricol;

x_3 — producția globală pe o normă convențională (lei).

Datele grupate sînt prezentate în tabelul 8.15.

[23] Informațiile statistice prelucrate privesc 160 CAP, dintr-un număr total de 229 CAP existente în județul Ilfov, în anul 1968. Volumul colectivității cercetate a fost redus în mod deliberat, în scopul de a se satisface întrucîtva condiția de omogenitate. Factorii aleși s-au testat cu ajutorul criteriului F.

Scanned with OKEN Scanner

Tabelul 8.15

x \ y	100–140	140–180	180–220	220–260	260–300
x_1					
100–140	13	7	2	1	
140–180	24	24	18	3	
180–220	4	16	15	6	2
220–260		3	6	4	7
260–300			2	2	1
x_2					
120–160	8	7	2	1	
160–200	19	13	4	4	
200–240	8	12	18	4	
240–280	3	13	13	2	2
280–320	2	4	3	3	3
320–360	1	1	3	2	5
x_3					
30–38	10	6	3	1	
38–46	17	20	17	3	
46–54	5	18	10	3	6
54–62	8	5	10	8	3
62–70	1	1	3	1	1

Ecuațiile normale:

$$\begin{aligned}
&\text{I } na + b_1\Sigma x_1 + b_2\Sigma x_2 + b_3\Sigma x_3 = \Sigma y \\
&\text{II } a\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2 + b_3\Sigma x_1x_3 = \Sigma x_1y \\
&\text{III } a\Sigma x_2 + b_1\Sigma x_1x_2 + b_2\Sigma x_2^2 + b_3\Sigma x_2x_3 = \Sigma x_2y \\
&\text{IV } a\Sigma x_3 + b_1\Sigma x_1x_3 + b_2\Sigma x_2x_3 + b_3\Sigma x_3^2 = \Sigma x_3y.
\end{aligned} \tag{8.28}$$

Calculele au condus la ecuațiile *:

$$\begin{aligned}
&\text{I } 160\,a + 28678\,b_1 + 36314\,b_2 + 7613\,b_3 = 28278 \\
&\text{II } 28678\,a + 5371658\,b_1 + 6726222\,b_2 + 1392094\,b_3 = 5234551 \\
&\text{III } 36314\,a + 6726222\,b_1 + 8866960\,b_2 + 1747675\,b_3 = 6643605 \\
&\text{IV } 7613\,a + 1392094\,b_1 + 1747675\,b_2 + 373489\,b_3 = 1369132.
\end{aligned} \tag{8.29}$$

* Pe baza datelor negrupate.

Scanned with OKEN Scanner

Acest sistem de ecuații normale poate fi rezolvat prin folosirea ecuației matriceale:

$$\hat{a} = (X^*X)^{-1}\ X^*Y, \tag{8.30}$$

unde $\hat{a}$ reprezintă vectorul coloană de parametri căutați**;

X — matricea valorilor variabilelor explicative observate empiric;

Y — vectorul coloană de valori empirice ale variabilei explicate.

În cazul sistemului (8.29), înlocuind în ecuația matriceală coeficienții parametrilor a, b_1, b_2, b_3

$$\begin{Vmatrix} 160 & 28768 & 36314 & 7613 \\ 28768 & 5371658 & 6726222 & 1392094 \\ 36314 & 6726222 & 8866960 & 1747675 \\ 7613 & 1392094 & 1747675 & 373489 \end{Vmatrix}^{-1} \begin{Vmatrix} 28278 \\ 5234551 \\ 6643605 \\ 1369132 \end{Vmatrix} = \begin{Vmatrix} a \\ b_1 \\ b_2 \\ b_3 \end{Vmatrix}$$

și efectuînd operațiile algebrice în matrice obținem:

$$a = 21{,}74788,\ b_1 = 0{,}48037,\ b_2 = 0{,}17425,\ b_3 = 0{,}61665.$$

Ecuația de regresie se scrie:

$$y^* = 21{,}74788 + 0{,}48037x_1 + 0{,}17425\ x_2 + 0{,}61665\ x_3. \tag{8.31}$$

Ca și în cazul regresiei simple, coeficienții regresiei multiple arată cu cît anume se schimbă variabila explicată, atunci cînd variabila explicativă respectivă se schimbă cu o unitate. Mărimea fiecărui coeficient de regresie este o mărime „netă", curățată de influențele exercitate în mod indirect de celelalte variabile explicative specificate în model. Astfel, din analiza regresiei (8.31) rezultă că venitul net realizat de către C.A.P. la 100 ha teren agricol, în anul pentru care s-au cules informațiile, a sporit, în medie cu:

a) 480 lei la un spor de cheltuieli materiale la 100 ha teren agricol (x_1) de 1000 lei (x_2 și x_3 menținuți la nivele constante);

b) 174 lei la un spor al valorii mijloacelor fixe pe 100 ha teren agricol (x_2) de 1000 lei (x_1 și x_3 menținuți la nivele constante);

c) 62 lei la un spor de producție globală pe o normă convențională (x_3) de 100 lei (x_1 și x_2 menținuți la nivele constante).

Funcția putere multiplă (8.24) este uzuală în modelarea multor fenomene, mai ales economice.

Exemplul 7. Pentru corelația dintre producție (y), fonduri fixe (x_1) și număr de personal (x_2) în industrie se consideră adecvat modelul:

$$y^* = ax_1^{b_1}x_2^{b_2} \tag{8.32}$$

logaritmizat

$$\log y^* = \log a + b_1 \log x_1 + b_2 \log x_2. \tag{8.33}$$

** Parametrii scriși a_0, a_1, a_2, ..., a_n.

Scanned with OKEN Scanner

Un sistem de trei ecuații normale

$$\begin{aligned}
&\text{I } n \log a + b_1 \Sigma \log x_1 + b_2 \Sigma \log x_2 = \Sigma \log y \\
&\text{II } \log a \Sigma \log x_1 + b_1 \Sigma (\log x_1)^2 + b_2 \Sigma (\log x_1 \log x_2) = \\
&\qquad = \Sigma (\log x_1 \log y) \\
&\text{III } \log a \Sigma \log x_2 + b_1 \Sigma (\log x_1 \log x_2) + b_2 \Sigma (\log x_2)^2 = \\
&\qquad = \Sigma (\log x_2 \log y).
\end{aligned} \tag{8.34}$$

Tabelul 8.16

Județul	Producția globală industrială	Fonduri fixe*	Lucrători	Județul	Producția globală industrială	Fonduri fixe *	Lucrători
Alba	1,4	1,3	2,2	Hunedoara	5,0	7,5	4,1
Arad	2,1	1,1	2,7	Ialomița	1,0	0,9	0,7
Argeș	3,6	4,8	3,2	Iași	2,7	2,0	2,1
Bacău	3,2	5,3	2,9	Ilfov	1,4	1,5	1,5
Bihor	2,4	2,2	3,1	Maramureș	1,5	1,6	2,3
Bistrița-Năsăud	0,3	0,2	0,5	Mehedinți	0,9	1,3	0,8
Botoșani	0,6	0,4	0,6	Mureș	3,4	3,3	3,2
Brașov	6,7	4,5	5,7	Neamț	2,5	3,4	2,1
Brăila	2,1	2,5	1,6	Olt	1,2	2,2	0,8
Buzău	1,2	1,4	1,2	Prahova	6,1	6,3	5,5
Caraș-Severin	2,5	2,6	2,6	Satu Mare	1,2	0,6	1,6
Cluj	3,5	2,9	4,3	Sălaj	0,2	0,3	0,5
Constanța	2,2	2,2	1,9	Sibiu	3,6	2,3	4,0
Covasna	0,6	0,5	0,9	Suceava	1,9	1,7	2,6
Dîmbovița	1,5	2,4	2,1	Teleorman	1,0	1,8	0,8
Dolj	3,3	4,0	2,3	Timiș	3,7	2,0	4,2
Galați	3,4	5,8	2,3	Tulcea	0,4	1,0	0,9
Gorj	1,0	2,3	1,4	Vaslui	0,7	0,5	0,7
Harghita	1,0	0,8	1,8	Vîlcea	1,0	1,7	1,0
				Vrancea	0,7	0,5	0,8

* Din unitățile socialiste
Sursa: *Anuarul Statistic al Republicii Socialiste România 1972*, p. 92—93.

Cu datele înregistrate*:

$$\begin{aligned}
&\text{I } 39a + 47{,}78129\, b_1 + 48{,}31821\, b_2 = 46{,}95785 \\
&\text{II } 47{,}78129\, a + 63{,}90881\, b_1 + 62{,}49792\, b_2 = 61{,}98051 \\
&\text{III } 48{,}31821\, a + 62{,}49792\, b_1 + 63{,}20596\, b_2 = 61{,}83695.
\end{aligned}$$

Sistemul se poate rezolva pe calea excluderii succesive a necunoscutelor (metoda Gauss):

a) excluderea necunoscutei $\log a$ prin împărțirea fiecărei ecuații inițiale la coeficientul de pe lîngă $\log a$:

$$\begin{aligned}
&\text{I' } 1{,}2251612\, b_1 + 1{,}2389284 b_2 = 1{,}2040474 \\
&\text{II' } 1{,}3375279\, b_1 + 1{,}30799986\, b_2 = 1{,}2971711 \\
&\text{III' } 1{,}2934651\, b_1 + 1{,}3081188\, b_2 = 1{,}2797856;
\end{aligned}$$

* S-a operat cu proporții la 1000.

b) excluderea necunoscutei b_1 prin scăderea din ecuaţiile II' şi III' a ecuaţiei I':

$$\text{I''}\quad 0{,}1123667\ b_1 + 0{,}0690714 b_2 = 0{,}0931237$$

$$\text{II''}\quad 0{,}0683039\ b_1 + 0{,}0691904\ b_2 = 0{,}0757382;$$

şi apoi împărţirea ecuaţiilor obţinute la coeficientul lui b_1:

$$\text{I'''}\quad 0{,}6146963\ b_2 = 0{,}8287481$$

$$\text{II'''}\quad 1{,}0129787\ b_2 = 1{,}1088415;$$

c) determinarea necunoscutei b_2, scăzînd ecuaţia I''' din ecuaţia II''':

$$b_2 = \frac{0{,}2800934}{0{,}3982824} = 0{,}7032532;$$

şi apoi, prin introducerea în ecuaţiile corespunzătoare a valorilor calculate succesiv, rezultă

$$b_1 = \frac{0{,}0931237 - (0{,}0690714 \times 0{,}7032532)}{0{,}1123667} = 0{,}3964617,$$

$$\log a = \frac{46{,}95785 - (47{,}78129 \times 0{,}3964617) - (48{,}31821 \times 0{,}7032532)}{39} =$$

$$= -0{,}1529624.$$

Pentru verificare se înlocuiesc valorile parametrilor în una din ecuaţiile normale iniţiale; de exemplu, în ecuaţia III: $-(48{,}31821 \times 0{,}1529624) + (62{,}49792 \times 0{,}3964617) + (63{,}20596 \times 0{,}7032532) = 61{,}8369559$.

Ecuaţia de regresie este:

$$\log y^* = -0{,}15296 + 0{,}39646 x_1 + 0{,}70325\ x_2. \qquad (8.35)$$

O variantă de rezolvare a sistemului de ecuaţii normale considerat, urmînd acelaşi principiu, poate fi:

a) ecuaţia I înmulţită cu $\Sigma \log x_1/n = 47{,}78129/39$ şi scăzută din ecuaţia II, apoi ecuaţia II înmulţită cu $\Sigma \log x_2/\Sigma \log x_1 = 48{,}31821/47{,}78129$ şi scăzută din ecuaţia III elimină o ecuaţie, rezultînd:

$$\text{I'}\quad 5{,}3690275\ b_1 + 3{,}3003239 b_2 = 4{,}4495742$$

$$\text{II'}\quad -2{,}1290332 b_1 + 0{,}0057509 b_2 = -0{,}8400349.$$

b) ecuaţia I' înmulţită cu raportul $-2{,}1290332/5{,}3690275$ şi scăzută din ecuaţia II' elimină b_1, de unde:

$$b_2 = 0{,}7032529$$

şi prin calcule succesive corespunzătoare:

$$b_1 = 0{,}3964613;\quad \log a = -0{,}1529616.$$

Scanned with OKEN Scanner

Întrucît variabilele explicative se exprimă deseori în unităţi de măsură diferite şi diferă în raport cu dispersia lor, pentru ierarhizarea influenţelor, coeficienţii de regresie parţială trebuie să fie aduşi la o formă comparabilă. La aceasta se poate ajunge, dacă se corectează coeficienţii de regresie iniţiali cu raportul dintre abaterea pătratică medie a variabilei explicative respective (s_x) şi abaterea pătratică medie a variabilei explicate (s_y). Adică, în cazul unei regresii bifactoriale:

$$\beta_{yx_1 \cdot x_2} = b_{yx_1 \cdot x_2} \frac{s_{x_1}}{s_y} \quad \text{şi} \quad \beta_{yx_2 \cdot x_1} = b_{yx_2 \cdot x_1} \frac{s_{x_2}}{s_y}. \tag{8.36}$$

Coeficienţii β, care sînt coeficienţi normaţi (standardizaţi) ai regresiei multiple [24], sînt comparabili între ei, fiindcă variabilele explicative s-au exprimat în aceeaşi unitate de măsură, în abaterea pătratică medie a variabilei explicate (s_y).

În legătură cu analiza regresiei s-a pus problema reversibilităţii ecuaţiei de regresie: cunoaşterea formei analitice a legăturii lui y faţă de x nu ne permite să determinăm legătura lui x faţă de y? Dacă, de exemplu, în ecuaţia de regresie venit naţional (y)/ investiţii (x)[25]

$$y^* = 28{,}972 + 2{,}753\ x \tag{8.37}$$

coeficientul b arată o creştere a venitului naţional la un miliard lei investiţii cu 2,753 miliarde lei, de aici n-ar reieşi că la un miliard lei venit naţional investiţiile vor trebui să se mărească, în medie, cu 1/2,753 = 0,363 miliarde lei? În cazul dependenţei stohastice, „operaţiunea cunoscută în matematică sub numele de inversiunea funcţiilor este o imposibilitate“ [47, p. 286]. Această operaţie este posibilă numai în cazul legăturii rigide, caracterizată printr-o strictă proporţionalitate în variaţiile variabilelor corelate şi, ca urmare, liniile de regresie y în x şi x în y coincid, iar coeficienţii de regresie $b_{y/x}$ şi $b_{x/y}$ sînt mărimi inverse.

Dependenţa statistică se defineşte tocmai prin aceea că variaţiile variabilelor corelate nu se află în raport de strictă proporţionalitate, coeficienţii de corelaţie au valori subunitare, liniile de regresie y în x şi x în y nu se suprapun, iar coeficienţii de regresie nu sînt mărimi inverse. Contradicţia aparentă se poate uşor explica, avînd în vedere împrejurarea că ecuaţia de regresie y/x se construieşte pe baza datelor statistice grupate după variabila x, iar ecuaţia de regresie x/y pe baza datelor grupate după variabila y, ceea ce antrenează după sine şi schimbarea în raportul dintre variaţiile ambelor variabile [38].

Aşadar, nici în cazul corelaţiei venit naţional/investiţii, deşi avem de-a face cu un raport de interdependenţă, cu o legătură „cu reacţie inversă“ (*feedback*), nu poate fi vorba de reversibilitatea funcţiei de modelare. Într-adevăr, ecuaţia de regresie, în care y — investiţii şi x — venit naţional, are forma:

$$y^* = -9{,}895 + 0{,}357\ x. \tag{8.38}$$

Ecuaţia (8.38) diferă de ecuaţia (8.37); dualitatea ajustării este inevitabilă (vezi şi [58]).

[24] Numiţi şi „coeficienţii beta“, după litera grecească cu care obişnuit se notează.
[25] Calculată pentru ţara noastră conform seriilor de date pe anii 1950—1969 [90, p. 121 şi 135].

Scanned with OKEN Scanner

8.5 COEFICIENTUL DE CORELAȚIE. RAPORTUL DE CORELAȚIE

Analiza de regresie este strîns legată de indicatorii intensității legăturii, deoarece semnificația și aplicabilitatea practică a ecuației de regresie depinde de gradul de legătură: în cazul unei legături slabe, ecuația de regresie ori este puțin sigură, ori nu are sens. Pe de altă parte, indicatorii intensității legăturii se determină prin forma ei. Finalitatea cercetării corelației se realizează pe calea îmbinării indicatorilor de regresie cu indicatorii de corelație propriu-ziși.

În arsenalul metodelor statistice există diverse procedee de măsurare a gradului de corelație între două sau mai multe variabile, unele mai simple și mai puțin exacte, altele mai complicate și mai exacte. Printre acestea din urmă, de aplicație curentă — în cazul legăturilor între caracteristici măsurabile — sînt: coeficientul de corelație și raportul de corelație.

8.5.1 Coeficientul de corelație

Coeficientul de corelație (r) măsoară gradul de legătură liniară între variabile și este definit de formula (cazul *a* două variabile)

$$r_{yx} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n s_x s_y}, \tag{8.39}$$

care prin dezvoltare devine:

$$r_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}. \tag{8.40}$$

El oferă avantajul de a fi un indicator abstract, independent de unitățile de exprimare a datelor originale și deci comparabil cu oricare altul.

De fapt, coeficientul de corelație s-ar putea socoti tot un coeficient de regresie, exprimat în abaterile normate pentru ambele variabile. Produsul mediu al abaterilor normate dă coeficientul de corelație:

$$\overline{t_x t_y} = r_{yx}, \tag{8.41}$$

în care

$$t_x = \frac{x - \bar{x}}{s_x} \text{ și } t_y = \frac{y - \bar{y}}{s_y}.$$

Coeficientul de regresie a lui y în x și coeficientul de corelație sînt legați prin egalitatea

$$b_{y/x} = r_{yx} \frac{s_y}{s_x}, \tag{8.42}$$

de unde

$$r_{yx} = b_{y/x} \frac{s_x}{s_y}. \tag{8.43}$$

Scanned with OKEN Scanner

Într-adevăr, dacă scriem

$$b_{y/x} = \frac{n\sum yx - \sum x \sum y}{n\sum x^2 - (\sum x)^2}; \quad s_x^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n} \quad \text{și} \quad s_y^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n},$$

atunci

$$\frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \cdot \frac{\sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}}}{\sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n}}} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \times \frac{\sqrt{n\sum x^2 - (\sum x)^2}}{\sqrt{n\sum y^2 - (\sum y)^2}} =$$

$$= \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = r_{yx}.$$

În relația (8.43), coeficientul de corelație apare drept un coeficient de regresie standardizat. Ecuația de estimare notată în forma:

$$y - \bar{y} = r \frac{s_y}{s_x} x - \bar{x} \tag{8.44}$$

arată că, în sens geometric, r este panta dreptei de regresie (estimare) în unități ale abaterii standard.

Semnificația coeficientului de corelație se desprinde cu claritate dintr-o altă formulare a sa:

$$r_{yx} = \sqrt{\frac{\sum (y^* - \bar{y})^2}{\sum (y - \bar{y})^2}}. \tag{8.45}$$

Adică, r_{yx} nu este altceva decît rădăcina pătrată din proporția variației totale a variabilei dependente, care a fost explicată de ecuația regresiei liniare simple.

În calculele de corelație, variația totală $\sum (y - \bar{y})^2$ se descompune în variația explicată $\sum (y^* - \bar{y})^2$ și variația neexplicată $\sum (y - y^*)^2$, în care y reprezintă valorile empirice ale variabilei dependente, iar y^*, valorile calculate ale aceleiași variabile. Între cele trei variabile există relația:

$$\sum (y^* - \bar{y})^2 + \sum (y - y^*)^2 = \sum (y - \bar{y})^2. \tag{8.46}$$

Coeficientul de corelație r este cuprins între -1 și $+1$:

$$-1 < r < +1.$$

El este pozitiv ($r > 0$) dacă ambele variabile x și y variază în același sens și este negativ ($r < 0$) în caz contrar. Cînd valoarea absolută a coeficientului de corelație este apropiată de unitate, se poate afirma că există o puternică dependență liniară între cele două variabile x și y. Dacă valoarea abso-

Scanned with OKEN Scanner

lută a coeficientului de corelaţie este slabă, nu trebuie în mod automat să se conchidă că nu există legătură între x şi y; legătura poate să existe, dar nu de formă liniară.

Exemplul 8. Să calculăm coeficientul de corelaţie între cerere şi venit (tabelul 8.4), aplicînd formula (8.40)

Elementele de calcul

$$n = 14 \qquad \Sigma x^2 = 576188{,}65$$
$$\Sigma x = 2695{,}9 \qquad \Sigma y^2 = 93595{,}87$$
$$\Sigma y = 1094{,}5 \qquad \Sigma xy = 232140{,}15,$$

introduse în formulă, ne dau:

$$r_{yx} = \frac{(14 \times 232140{,}15) - (2695{,}9 \times 1094{,}5)}{\sqrt{[(14 \times 576188{,}65) - (2695{,}9)^2]\,[(14 \times 93595{,}87) - (1094{,}5)^2]}} = 0{,}998.$$

Dispunînd de datele incluse în tabelul 8.4, putem folosi de asemenea formula (8.45). Într-adevăr:

$$r_{yx} = \sqrt{\frac{8010{,}81}{8029{,}41}} = 0{,}998.$$

O corelaţie pozitivă (variaţiile în acelaşi sens) foarte puternică există între cele două variabile.

Exemplul 9. În tabelul 8.2, căsătoriile sînt grupate pe intervale de vîrstă ale soţilor (tabel de corelaţie). Coeficientul de corelaţie se calculează după formula:

$$r_{yx} = \frac{\Sigma f \Sigma xy\, f_{xy} - (\Sigma x f_x)\,(\Sigma y f_y)}{\sqrt{[\Sigma f \Sigma x^2 f_x - (\Sigma x f_x)^2]\,[\Sigma f \Sigma y^2 f_y - (\Sigma y f_y)^2]}}. \tag{8.47}$$

Elementele de calcul se găsesc în tabelul 8.2. Introducîndu-se în formulă obţinem $r_{yx} = 0{,}822$.

Cum intervalele de grupare sînt egale, volumul operaţiilor de calcul poate fi redus, dacă se consideră valorile micşorate ale variabilelor corelate (tabelul 8.17):

$$x' = \frac{x - 37}{5} \quad \text{şi} \quad y' = \frac{y - 37}{5}.$$

Astfel, pe baza elementelor de calcul din tabelul 8.17,

$$r_{yx} = \frac{(175496 \times 1458994) - (-384354)\,(-509243)}{\sqrt{[(175496 \times 1280068) - (-384354)^2]\,[(175496 \times 1876511) - (-509243)^2]}} =$$
$$= 0{,}822.$$

În masă, căsătoriile se caracterizează printr-o strînsă legătură între vîrstele soţilor.

Tabelul 8.17

y	y' \ x	x'	17	22	27	32	37	42	47	52	57	62	f_y	$y'f_y$	$(y')^2f_y$	$\Sigma y'x'f_{xy}$
			−4	−3	−2	−1	0	1	2	3	4	5				
17	−4		6584	44833	12738	1311	179	26	11	5	3	4	65694	−262776	1051104	750108
22	−3		1870	44278	23956	4862	1141	215	69	25	9	12	76437	−229311	687933	577692
27	−2		112	3222	6671	3788	1543	433	147	46	7	7	15976	−31952	63904	52632
32	−1		21	367	1051	1672	1463	673	261	74	21	26	5629	−5629	5629	3328
37	0		3	105	299	658	1016	880	435	179	51	60	3686	0	0	0
42	1		0	28	78	179	439	616	594	336	104	181	2555	2555	2555	3714
47	2		1	10	9	37	129	255	477	461	226	380	1985	3970	7940	10614
52	3		0	2	2	7	22	49	182	283	270	696	1513	4539	13617	17415
57	4		0	7	6	1	1	10	28	59	125	507	744	2976	11904	12976
62	5		1	4	6	1	0	2	13	28	90	1132	1277	6385	31925	30515
f_x			8592	92856	44816	12516	5933	3159	2217	1496	906	3005	175496	−509243	1876511	1458994
$x'f_x$			−34368	−278568	−89632	−12516	0	3159	4434	4488	3624	15025	−384354			
$(x')^2f_x$			137472	−885704	179264	12516	0	3159	8868	13464	14496	75125	1280068			
$\Sigma x'y'f_x{}^y$			128736	956625	274114	28795	0	−965	2930	6666	8968	53125	1458994			

Scanned with OKEN Scanner

Pentru a măsura gradul de legătură liniară între o variabilă dependentă y și mai multe variabile independente $x_1, x_2, \ldots, x_n$ se utilizează *coeficientul corelației multiple* (R) dedus din formula:

$$R_{yx_1x_2\ldots x_n} = \sqrt{1 - \frac{\Sigma(y - y^*_{x_1x_2\ldots x_n})^2}{\Sigma(y - \bar{y})^2}}. \tag{8.48}$$

Valoarea lui R depinde de raportul dintre dispersia valorilor determinate pe baza ecuației de regresie liniară multiplă și dispersia valorilor observate ale variabilei dependente. Cu cît valorile observate se abat mai puțin de la linia de regresie multiplă cu atît coeficientul de corelație are o valoare mai mare și deci corelația este mai intensă. În ipoteza $\Sigma(y - y^*_{x_1x_2\ldots x_n})^2 = 0$, $R_{yx_1x_2\ldots x_n} = 1$, ceea ce ar însemna că valorile observate ale lui y coincid cu valorile calculate pe baza ecuației de regresie, variația variabilei y fiind integral explicată prin variația variabilelor independente incluse în calculul de corelație.

Dacă s-a rezolvat ecuația de regresie, coeficientul de corelație se calculează cu ușurință după formula:

$$R_{yx_1x_2\ldots x_n} = \sqrt{\frac{a\Sigma y + b_1\Sigma x_1y + b_2\Sigma x_2y + \ldots + b_n\Sigma x_ny - \frac{(\Sigma y)^2}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}. \tag{8.49}$$

Coeficientul corelației multiple măsoară influența simultană a variabilelor independente $x_1x_2, \ldots, x_n$ asupra variabilei dependente y. În analiza corelației prezintă interes însă și determinarea gradului de legătură dintre variabila dependentă și o singură variabilă independentă, cu excluderea influenței celorlalte variabile independente. În acest scop se recurge la *coeficienții corelației parțiale*, care se pot calcula, în cazul a trei variabile, cu ajutorul coeficienților corelației simple:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2}r^2_{x_1x_2}}{\sqrt{1 - r^2_{yx_2}}\sqrt{1 - r^2_{x_1x_2}}} \tag{8.50}$$

și

$$r_{yx_2 \cdot x_1} = \sqrt{\frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{1 - r^2_{yx_1}}\sqrt{1 - r^2_{x_1x_2}}}}, \tag{8.51}$$

iar în cazul a patru variabile din coeficienții corelației parțiale de ordinul întîi:

$$r_{yx_1 \cdot x_2x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2}r_{x_1x_3 \cdot x_2}}{\sqrt{1 - r^2_{yx_3 \cdot x_2}}\sqrt{1 - r^2_{x_1x_3 \cdot x_2}}} \tag{8.52}$$

$$r_{yx_2 \cdot x_1x_3} = \frac{r_{yx_2 \cdot x_1} - r_{yx_3 \cdot x_1}r_{x_2x_3 \cdot x_1}}{\sqrt{1 - r^2_{yx_3 \cdot x_1}}\sqrt{1 - r^2_{x_2x_3 \cdot x_1}}} \tag{8.53}$$

$$r_{yx_3 \cdot x_1x_2} = \frac{r_{yx_3 \cdot x_1} - r_{yx_2 \cdot x_1}r_{x_2x_3 \cdot x_1}}{\sqrt{1 - r^2_{yx_2 \cdot x_1}}\sqrt{1 - r^2_{x_2x_3 \cdot x_1}}}. \tag{8.54}$$

Scanned with OKEN Scanner

Să adăugăm că şi coeficienţii corelaţiei multiple pot fi obţinuţi, de exemplu:

— în cazul corelaţiei de trei variabile din coeficienţii de ordinul zero [26]

$$R_{yx_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}}}; \tag{8.55}$$

— în cazul corelaţiei de patru variabile din relaţia

$$R_{yx_1x_2x_3} = \sqrt{R^2_{yx_1x_2} + r^2_{yx_3 \cdot x_1x_2}(1 - R^2_{yx_1x_2})} \tag{8.56}$$

sau

$$R_{yx_1x_2x_3} = \sqrt{r^2_{yx_1} + r^2_{yx_2 \cdot x_1}(1 - r^2_{yx_1}) + r^2_{yx_3 \cdot x_1x_2}(1 - R^2_{yx_1x_2})} \tag{8.57}$$

Utilitatea coeficienţilor de corelaţie parţială poate fi văzută din aplicaţia la exemplul 7. Intensitatea corelaţiei dintre producţie (y), fonduri fixe (x_1) şi număr de lucrători (x_2) este exprimată de următorii coeficienţi:

$$R_{\log y \log x_1 \log x_2} = 0{,}957$$

$$r_{\log y \log x_1 \cdot \log x_2} = 0{,}676$$

$$r_{\log y \log x_2 \cdot \log x_1} = 0{,}789.$$

Conform coeficientului de corelaţie multiplă, mărimea producţiei industriale se află în dependenţă foarte strînsă faţă de volumul fondurilor fixe şi de numărul de lucrători. Dependenţa apare mai intensă faţă de numărul de lucrători decît faţă de volumul fondurilor fixe, aşa cum rezultă din calculul corelaţiei parţiale.

8.5.2 Raportul de corelaţie

Corelaţia neliniară este descrisă sub aspectul intensităţii de *raportul de corelaţie* (η), cunoscut şi sub denumirea de indice de corelaţie curbilinie, definit de formula:

$$\eta_{yx} = \sqrt{1 - \frac{\Sigma(y - y^*_x)^2}{\Sigma(y - \bar{y})^2}} \tag{8.58}$$

[26] Sînt denumiţi coeficienţi de ordinul zero — coeficienţii corelaţiei de două variabile, coeficienţi de ordinul întîi — coeficienţii corelaţiei parţiale de trei variabile, coeficienţi de ordinul doi — coeficienţii corelaţiei parţiale de patru variabile ş.a.m.d. În general, ordinul indică numărul variabilelor considerate, statistic, constante (în cazul coeficientului de ordinul zero, nici o variabilă nu se consideră constantă). De menţionat că în scrierea coeficienţilor, de pildă $r_{yx_1 \cdot x_2}$ sau $r_{yx_1 \cdot x_2x_3}$, semnele din dreapta punctului la indice arată variabilele care au fost considerate constante.

Scanned with OKEN Scanner

sau, ceea ce este totuna:

$$\eta_{yx} = \sqrt{\frac{\Sigma (y_x^* - \bar{y})^2}{\Sigma (y - \bar{y})^2}}. \tag{8.59}$$

Raportul de corelație are la bază descompunerea dispersiei totale a variabilei dependente în dispersia valorilor empirice față de valorile teoretice (estimate printr-o ecuație de regresie neliniară) și dispersia valorilor teoretice de la medie. Compararea dispersiei valorilor teoretice de la medie cu dispersia valorilor empirice de la medie dă raportul de corelație [27].

Ca ilustrare să aplicăm formula (8.58) la calculul raportului de corelație, care să exprime gradul de legătură parabolică între producția de porumb și densitatea plantelor (exemplul 4).

Rezolvarea sistemului de ecuații normale (8.17) prin metoda simplex ne oferă elementele de calcul necesare (vezi tabelele 8.10 și 8.12)

$$\Sigma(y - y^*)^2 = 1{,}33640 \text{ și } \Sigma(y - \bar{y})^2 = 97{,}52.$$

$$\eta_{yx\,x^2} = \sqrt{1 - \frac{1{,}3364}{97{,}52}} = \sqrt{\frac{96{,}1836}{97{,}52}} = 0{,}993.$$

Deoarece s-a rezolvat ecuația de regresie, calculul se poate efectua și după formula:

$$\eta_{yxx^2} = \sqrt{\frac{a\Sigma y + b\Sigma xy + b'\Sigma x^2 y - \frac{(\Sigma y)^2}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}.$$

Elementele de calcul se găsesc:

— în ecuația de regresie (8.18)

$$a = 25{,}217;\ b = 1{,}0637;\ b' = -0{,}00975$$

[27] K. Pearson a propus un raport de corelație, care nu implică nici o ipoteză *apriori* asupra formei legăturii:

$$\eta_{yx} = \sqrt{1 - \frac{C_{yx}^2}{s_y^2}}, \tag{8.60}$$

în care C_{yx}^2 reprezintă media pătratelor abaterilor de la curba de regresie y în x (curbă descrisă de mediile condiționate $\hat{y}_x$). Semnificația lui η_{yx} apare mai clară din calculul său sub forma transformată:

$$\eta_{yx} = \sqrt{\frac{\Sigma(\hat{y}_x - \bar{y})^2 f_i}{\Sigma(y - \bar{y})^2 f_j}}. \tag{8.61}$$

Vezi aplicație (109, p. 380–381).

Scanned with OKEN Scanner

— ecuaţiile normale (8.17′);

$\Sigma y = 303$; $\Sigma xy = 13961$; $\Sigma x^2 y = 727650$;

— în tabelul 8.9:

$\Sigma y^2 = 15399{,}02$,

de unde

$$\eta_{yxx^2} = \sqrt{\frac{(25{,}217 \times 303) + (1{,}0637 \times 13961) - (0{,}00975 \times 727650) - \frac{(303)^2}{6}}{15399{,}02 - \frac{(303)^2}{6}}} = 0{,}993.$$

După formule asemănătoare se calculează şi raportul de corelaţie pentru legături neliniare multiple.

8.6 CORELAŢII TEMPORALE

În materie de comparaţie a seriilor cronologice se face apel la calculele de corelaţie în condiţii particulare. De fapt, trei cazuri sînt de observat:

a) o corelaţie între valorile luate de o aceeaşi variabilă statistică la două momente separate printr-un anumit interval de timp *(corelaţie internă* sau *autocorelaţie)*;

b) o corelaţie între valorile luate de două variabile statistice distincte (x şi y) la un acelaşi moment *(corelaţie sincronă);*

c) o corelaţie între valoarea luată de o variabilă statistică (x) la un anumit moment şi valoarea unei alte variabile (y), după un interval de timp determinat *(corelaţie asincronă* sau *corelaţie cu întîrziere* sau *corelaţie cu decalaje de timp).*

Cel de al doilea caz a fost tratat, în mod implicit, în paragrafele precedente. Cît priveşte primul caz, în care se cere eliminarea influenţei trendului, cea mai bună soluţie este modelul în care participă timpul (t) ca una din variabilele independente [28]:

$$y_{(t)} = f(x_1, x_2, \ldots, x_n; t) + \varepsilon_{(t)} \qquad (8.62)$$

[28] Căutînd să stabilească pentru Anglia corelaţia între nupţialitate şi comerţul exterior, R. H. Hooker a sesizat, se pare, cel dintîi dificultatea provocată de existenţa trendurilor şi a utilizat „un mijloc foarte simplu de a depăşi această dificultate, considerînd abaterile celor două variabile de la media instantanee (medie centrată. — *n.n.*) şi nu de la media ansamblului perioadei...“ [133]. R. H. Hooker a expus metoda sa, cu aplicaţie la corelaţia menţionată, în memoriul *On the correlation of the mariage — rate with trade,* „Journal of the Royal Statistical Society“, 1901. Alte procedee de eliminare a trendului, de mult elaborate, sînt prezentate în capitolul consacrat corelaţiei seriilor în timp, de Arthur L. Bowley, în *Element of Statistics,* ediţia a 5-a, Londra, 1926. Vezi şi [90].

Scanned with OKEN Scanner

sau circumscris la raportul y/x de formă liniară:

$$y_{(t)} = a + b_1 x + b_2 t + \varepsilon_{(t)}. \tag{8.63}$$

Într-un astfel de model, se exprimă dependența variabilei y nu numai de factorii specificați $x_1, x_2, \ldots, x_n$, ci formal și de t, luat drept exponent general al factorilor reziduali. Este o cale comodă pentru determinarea corelației parțiale, făcînd variabila t constantă [29].

Exemplul 10 Rezolvarea modelului 8.63 cu datele tabelului 8.18, privind corelația dintre venitul național pe o persoană ocupată (x) și înzestrarea tehnică a muncii (y) în industria țării noastre, certifică necesitatea luării în considerare a influenței trendurilor.

Tabelul 8.18
1950 = 100

Anii	Venitul național pe o persoană ocupată	Fonduri fixe pe o persoană ocupată	Anii	Venitul național pe o persoană ocupată	Fonduri fixe pe o persoană ocupată
1951	123	104	1963	321	178
1952	142	112	1964	353	186
1953	154	116	1965	381	192
1954	154	116	1966	405	209
1955	178	125	1967	448	233
1956	178	133	1968	489	247
1957	199	143	1969	518	271
1958	219	147	1970	568*	291
1959	236	153	1971	600*	306
1960	259	156	1972	630*	319
1961	276	158	1973	661*	333
1962	298	162	1974	729*	369

* Estimații pe baza ecuației de regresie (8.64)

Sursa: *Anuarul statistic al Republicii Socialiste România 1966*, p. 104, 106; *1970*, p. 112–113, 116–117; *1975*, p. 57–58, 68.

În corelația cercetată, se poate admite că variabila independentă (x) își exercită influența asupra variabilei dependente (y) în cea mai mare măsură în același an și deci avem de-a face cu o *corelație sincronă.*

[29] Din punct de vedere matematic, corelația parțială cu excluderea variabilei t (modelul 8.63) este exact echivalentă cu corelația dintre abaterile valorilor empirice de la trendurile lor liniare.

Scanned with OKEN Scanner

Ecuaţiile de regresie şi coeficienţii de corelaţie în variantele de calcul cu (1) şi fără (2) luarea în considerare a trendurilor sînt:

Ecuaţiile de regresie:

$$(1)\quad y^* = -79{,}87502 + 1{,}67483\,x + 7{,}62084\,t \tag{8.64}$$

$$(2)\quad y^* = -144{,}55388 + 2{,}56775\,x \tag{8.65}$$

Coeficienţii de corelaţie:

(1) $r_{yx \cdot t} = 0{,}889$

(2) $r_{yx} = 0{,}992.$

Coeficienţii β

$\beta_{yx \cdot t} = 0{,}64660$

$\beta_{yt \cdot x} = 0{,}35749.$

Diminuarea accentuată a corelaţiei în urma eliminării influenţei trendurilor era de aşteptat, dată fiind existenţa unei autocorelaţii puternice în seriile cronologice ale variabilelor corelate. Coeficienţii de autocorelaţie au valorile:

$$r_{y_i\, y_{i+1}} = 0{,}998 \quad \text{şi} \quad r_{x_i\, x_{i+1}} = 0{,}950.$$

Reacţiile între variabile nu sînt în mod necesar imediate; ele implică adesea un anumit decalaj în timp. Pentru decalajul de o singură perioadă, un model unifactorial are forma:

$$y_{(t)} = f(x_{(t-1)}) + \varepsilon_{(t)}; \tag{8.66}$$

în cazul funcţiei liniare:

$$y_{(t)} = a + bx_{(t-1)} + \varepsilon_{(t)} \tag{8.67}$$

Unei valori y_i a variabilei y, observată la timpul t_i, i se asociază o valoare x_{i-1} a variabilei x observată la timpul t_{i-1} şi se calculează ecuaţia de regresie şi coeficientul de corelaţie între cele două serii astfel decalate.

Exemplul 11. Să ne referim la o corelaţie demo-economică şi de domeniul istoriei însă tipică şi instructivă pentru cazul general, şi anume — cu titulatura din documente — raporturile între nupţialitate şi producţia de grîu şi porumb. Este corelaţia de care s-au ocupat primele studii analitice privind mişcarea populaţiei ţării noastre. În prefaţa la volumul *Mişcarea populaţiei României în anii 1901—1903*, dr. Leonida Golescu scria: „În prima linie, vorbind despre căsătorii, trebuie să reamintim că ele sînt direct influenţate de starea economică a ţării şi, în special la noi, de rezultatele recoltelor“ [30].

[30] *Mişcarea populaţiei României în anii 1901—1903*, Bucureşti, 1911 p. 7.

Scanned with OKEN Scanner

În volumul următor (anii 1904—1905), dr. I. N. Angelescu încearcă să stabilească, pentru prima dată în statistica noastră, o corelație între producția de grîu și porumb și nupțialitate, construind seriile de date din tabelul 8.19.

Tabelul 8.19

Anii	Nupțialitatea (căsătorii la 1000 locuitori)	Producția de grîu la ha (hl)	Producția de porumb la ha (hl)
1899	16,8	5,5	4,8
1900	13,4	12,6	14,7
1901	14,3	15,6	12,4
1902	17,9	18,1	11,1
1903	17,4	16,1	13,7
1904	16,2	11,0	3,3
1905	15,8	18,5	10,6
1906	20,3	19,8	22,1
1907	21,0	8,7	10,5
1908	18,2	10,7	13,8
1909	18,4	11,8	11,6
1910	18,4	20,0	18,4
1911	21,0	17,1	18,7
1912	17,2	11,1	17,6

Sursa: *Mișcarea populației României în anii 1904 și 1905*, *București* 1915.

Pentru a pune în evidență mai clar „mersul nupțialității în raport cu producția de grîu și porumb", seriile statistice au fost reprezentate grafic, de unde s-a desprins observația: „ ... efectele relei producții se simt mai puternic în cursul anului următor, pînă la noua producție". Și mai departe: „Așadar, în tot acest interval de 14 ani, nupțialitatea variază, în mod foarte regulat, după producția de grîu și porumb. Atît de puternică e influența producției acestor cereale asupra nupțialității încît se pare că ar fi singurele cauze ale ei". Analiza corelației s-a oprit aici.

Dacă vom calcula acum coeficienții de corelație pereche, fără a lua în considerare efectele decalate, vom găsi $r_{yx_1} = 0{,}112$, și $r_{yx_1} = 0{,}366$ (x_1 — producția de grîu la ha și x_2 — producția de porumb la ha). În consecință, variația producției de grîu n-ar explica decît 1% ($r^2 = 0{,}112^2$) din variația nupțialității, iar variația producției de porumb, circa 13% ($r^2 = 0{,}366^2$).

Efectuînd apoi calculele de corelație cu decalajul de un an sugerat de diagramă, rezultă următorii coeficienți de corelație: $r_{yx_1} = 0{,}668$ și $r_{yx_2} = 0{,}573$. Cum era de așteptat, influențele celor două variabile asupra nupțialității, prinse în anii următori recoltei, cînd produsele veneau „în consumație și vînzare", se înscriau cu ponderi importante, 44,6% și 32,8%. Cu toate acestea, dînd crezare calculelor de corelație, ipoteza studiului analitic, potrivit căreia producția de grîu și porumb „ar fi singurele cauze ale nupțialității", este infirmată. Coeficientul corelației multiple are valoarea 0,686, ceea ce înseamnă un grad de influență comună egal cu numai 47%. Nu se confirmă nici influența mai puternică a producției de porumb decît a producției de grîu presupusă de același studiu. [31]

[31] *Mișcarea populației României în anii 1904—1905*, p. 6.

În practică, decalajul de timp, cu care trebuie să se opereze în măsurarea corelației, nu se conturează atît de clar ca în seriile cronologice din aplicația citată. De aceea, se recomandă calculul coeficienților de corelație cu decalaje de una, două, trei etc. perioade și se reține numărul de perioade decalate pentru care există corelația cea mai strînsă.

8.7 MĂSURAREA INFLUENȚEI FACTORILOR ÎN RELAȚIA STOHASTICĂ

Calculele de corelație mijlocesc nu numai estimarea unei variabile în funcție de alta sau altele, măsurarea gradului de legătură între variabile, ci și evaluarea influenței factorilor. Judecățile despre intensitatea influenței comune și separate a factorilor cercetați se fac după mărimea *coeficienților de determinare.*

Într-un model unifactorial liniar, pătratul coeficientului de corelație (r^2) este coeficientul de determinare. În exemplul 2, s-a găsit un coeficient de corelație între cerere și venit egal cu 0,998, deci un coeficient de determinare în mărime de $(0{,}998)^2 = 0{,}996$. Aceasta înseamnă că variațiile cererii de mărfuri se explică în proporție de 99,6% prin variațiile veniturilor bănești ale populației.

Cînd modelul este multifactorial, măsurarea influenței factorilor constă în descompunerea variației variabilei dependente pe seama fiecărei variabile independente. Ca unitate de măsură a influenței simultane, exercitată de variația tuturor factorilor specificați în model, se indică pătratul coeficientului de corelație multiplă, numit *coeficient de determinare multiplă* (R^2). Semnificația lui R^2 rezultă din formula de calcul scrisă în felul următor:

$$D_{yx_1x_2\dots x_n} = R^2_{yx_1x_2\dots x_n} = \frac{\frac{\Sigma(y-\bar{y})^2}{n} - \frac{(\Sigma y - y^*_{x_1x_2\dots x_n})^2}{n}}{\frac{\Sigma(y-\bar{y})^2}{n}}. \tag{8.68}$$

Numărătorul reprezintă mărimea cu care se micșorează dispersia variabilei y datorită eliminării influenței factorilor $x_1, x_2, \dots, x_n$; raportată la $\Sigma(y-y)^2/n$, exprimă însăși ponderea influenței simultane a variației factorilor considerați.

Stabilirea influenței separate a factorilor se poate realiza prin descompunerea determinării totale. Astfel, formula de calcul a coeficientului determinării multiple (cazul a două variabile independente), care se poate scrie:

$$D_{yx_1x_2} = 1 - \frac{\frac{\Sigma(y-\bar{y})^2}{n} - b_{yx_1.x_2}\left(\frac{\Sigma xy_1}{n} - \bar{y}\bar{x}_1\right) - b_{yx_2.x_1}\left(\frac{\Sigma yx_2}{n} - \bar{y}\bar{x}_2\right)}{\frac{\Sigma(y-\bar{y})^2}{n}} \tag{8.69}$$

Scanned with OKEN Scanner

se transformă în

$$D_{yx_1x_2} = \frac{b_{yx_1.x_2}\left(\frac{\Sigma yx_1}{n} - \bar{y}\bar{x}_1\right) + b_{yx_2.x_1}\left(\frac{\Sigma yx_2}{n} - \bar{y}\bar{x}_2\right)}{\frac{\Sigma(y-\bar{y})^2}{n}}$$

și, divizată, partea dreaptă a expresiei ne dă determinarea parțială:

$$D_{yx_1x_2} = \frac{b_{yx_1.x_2}\left(\frac{\Sigma yx_1}{n} - \bar{y}\bar{x}_1\right)}{\frac{\Sigma(y-\bar{y})^2}{n}} + \frac{b_{yx_2.x_1}\left(\frac{\Sigma yx_2}{n} - \bar{y}\bar{x}_2\right)}{\frac{\Sigma(y-\bar{y})^2}{n}}. \qquad (8.70)$$

Formule mai convenabile pentru calcule sînt:

$$d_{yx_1 \cdot x_2} = b_{yx_1 \cdot x_2}\frac{n\Sigma x_1 y - \Sigma x_1 \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}; \quad d_{yx_2 \cdot x_1} = b_{yx_2 \cdot x_1}\frac{n\Sigma x_2 y - \Sigma x_2 \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}. \qquad (8.71)$$

Exemplul 12. Prezentăm rezultatele aplicării metodei corelației la determinarea influenței volumului fondurilor fixe productive (x_1) și a numărului populației ocupate (x_2) asupra venitului național din industrie (y). Seriile de date inițiale figurează în tabelul 8.20.

Tabelul 8.20

procente (1950=100)

Anii	Venitul național (prețuri comparabile)	Fonduri fixe productive (la sfîrșitul anului)	Populația ocupată (la sfîrșitul anului)	Anii	Venitul național (prețuri comparabile)	Fonduri fixe productive (la sfîrșitul anului)	Populația ocupată (la sfîrșitul anului)
1951	129	109	105	1961	424	243	154
1952	151	118	106	1962	484	263	162
1953	172	130	112	1963	540	299	168
1954	183	138	119	1964	620	327	176
1955	218	153	122	1965	710	357	186
1956	224	167	126	1966	782	404	193
1957	253	181	127	1967	884	460	197
1958	287	192	131	1968	994	303	203
1959	318	206	135	1969	1100	586	216
1960	372	225	144	1970	1300	664	228

Sursa: Anuarul statistic al Republicii Socialiste România 1966, p. 104, 106; *1971*, p. 111, 115, 125; *1972*, p. 90—91, 99.

Dinamica venitului național în raport cu dinamica celor doi factori s-a descris prin funcția de forma

$$y^*_{yx_1x_2} = 1{,}1839\, x_1^{0{,}47262}\, x_2^{1{,}90339} \qquad (8.72)$$

Coeficientul corelației multiple are valoarea:

$$R_{\log y \log x_1 \log x_2} = 0{,}99701.$$

Cu elementele de calcul

$\sum \log y = 12{,}09087$

$\sum \log x_1 = 3{,}58504$

$\sum \log x_2 = 7{,}88681$

$\sum \log x_1 \log y = 2{,}79163$

$\sum \log x_2 \log y = 6{,}15355$

$\sum (\log y)^2 = 9{,}12123$

$n = 21$

și valorile coeficienților de regresie din ecuația (8.72) introduse în formula (8.71) se obțin:

$$d_{yx_1 \cdot x_2} = 0{,}35289 \text{ și } d_{yx_2 \cdot x_1} = 0{,}64114.$$

Reiese că din dispersia valorilor venitului național, în intervalul considerat, circa 35% revin variației volumului fondurilor fixe și 64% variației numărului persoanelor ocupate; influența întrunită $(0{,}997)^2 = 0{,}994$ sau 99%.

În 8.4.2. am rezolvat același model pentru seriile de date pe județe, avînd expresia:

$$y^* = 0{,}70314\, x_1^{0{,}39646}\, x_2^{0{,}70325}.$$

Calculul coeficienților de determinare parțială evidențiază o ierarhizare a influențelor apropiată de cea indicată de modelul dinamic:

$$d_{yx_1 \cdot x_2} = 0{,}37271$$

$$d_{yx_2 \cdot x_1} = 0{,}54374$$

8.8 PREVIZIUNEA PE BAZA FUNCȚIEI DE MODELARE A CORELAȚIEI

Cînd se acceptă ipoteza că procesul de formare a unui fenomen în trecut se va prelungi în viitor, cel puțin pe termen scurt, fără modificări sensibile, metodele statistice, în special analiza de corelație, permit să se facă previziuni cu toată obiectivitatea. Prin intermediul ecuației de regresie se estimează valorile cele mai probabile ale funcției, corespunzător anumitor valori date argumentelor.

Dar extrapolarea valabilității în viitor a ecuației de regresie necesită o atentă verificare prealabilă. Dealtfel, operația de verificare a semnificației indicatorilor de corelație calculați este obligatorie indiferent de destinația lor: analiză, generalizări teoretice sau previziune.

Să considerăm ecuația de regresie:

$$y_x^* = 6{,}0241173 + 0{,}3747032\, x,$$

în care y reprezintă vînzările de mărfuri și x — veniturile bănești ale populației, rezolvată pe baza seriilor cronologice din tabelul (8.4).

Scanned with OKEN Scanner

Semnificaţia parametrului b se verifică cu ajutorul criteriului t a lui Student, definit de formula:

$$t_{calc.} = \frac{b\sqrt{\Sigma (x - \bar{x})^2 (n - 2)}}{\sqrt{\Sigma (y - y^*)^2}} \tag{8.73}$$

sau

$$t_{calc.} = \frac{b\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right](n-2)}}{\sqrt{\Sigma y^2 - a\Sigma y - b\Sigma xy}}, \tag{8.74}$$

formulă obţinută prin transformări elementare ale formulei (8.73), simplificînd calculele.

Se consideră b semnificativ, cînd valoarea lui t, obţinută din datele empirice, este mai mare decît valoarea sa tabelară, [32] corespunzătoare nivelului de semnificaţie P şi numărului gradelor de libertate $f = n - 2$.

Intervalul de încredere, în care se va afla — cu probabilitatea aleasă — coeficientul de regresie b, se exprimă prin formula [33].

$$b \pm t_{teor.} \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right](n-2)}}. \tag{8.75}$$

În cazul aplicaţiei, rezultatele calculelor de verificare sînt:

$s_b = 0{,}0629$

$t_{calc.} = \frac{b}{s_b} = 5{,}9562$

grade de libertate $= 14 - 2 = 12$

$t_{0,05} = 2{,}179$

intervalul de încredere $= b \pm t_{0,05}\, s_b = 0{,}2376 \div 0{,}5118$

Întrucît $t_{calc.} = 5{,}9562 > t_{teor.} = 2{,}179$, coeficientul de regresie este semnificativ, iar valorile lui se găsesc într-un interval de încredere îngust, garantat de probabilitatea 95%.

Verificarea statistică trebuie extinsă şi asupra coeficientului de corelaţie, ca măsurător al intensităţii legăturii. Semnificaţia coeficientului de corelaţie simplă se verifică cu testul t, calculat după formula:

$$t_{calc.} = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}} \tag{8.76}$$

[32] Vezi repartiţia t în anexa 5.

[33] Partea din expresia (8.75)

$$\sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right](n-2)}} = \sqrt{\frac{\Sigma (y - y^*)^2}{\Sigma (x - \bar{x})^2 (n-2)}} = s_b$$

reprezintă eroarea medie a parametrului b.

Se emite ipoteza $\rho = 0$, față de ipoteza alternativă $\rho > 0$, și se verifică prin confruntarea valorii efective a lui t cu valoarea sa din tabel, pentru nivelul de semnificație stabilit [34]. Dacă t calculat (8.76) este mai mare decît t teoretic, atunci coeficientul empiric de corelație se consideră semnificativ.

Calculul caracteristicii statistice a ipotezei cu formula (8.76), în cazul modelului de corelație dintre cerere și venit, dă:

$$t_{calc.} = \frac{0{,}99882\sqrt{14-2}}{\sqrt{1-(0{,}99882)^2}} = 71{,}22.$$

Din tabelul repartiției t cu $P = 0{,}05$ și $f = 12$ grade de libertate, aflăm $t = 2{,}179$. Comparația lui $t_{calc.}$ cu $t_{teor.}$ infirmă ipoteza nulă ($t_{calc.} > t_{teor.}$); coeficientul de corelație are valoare semnificativă.

Domeniul de încredere a lui t se calculează din

$$s_r = \frac{1 - r^2}{\sqrt{n-1}} \tag{8.77}$$

avînd limitele

$$r \pm t_{teor.}\, s_r \tag{8.78}$$

(s_r este eroarea pătratică medie)

În aplicație:

$$s_r = \frac{1 - (0{,}99882)^2}{\sqrt{14-1}} = 0{,}000654$$

și

$$0{,}99882 \pm 2{,}160 \times 0{,}000654.$$

Riscul acceptat pentru depășirea limitelor domeniului de încredere este de 5%.

Caracterul adecvat al ecuației de regresie se verifică direct prin compararea valorilor estimate cu valorile observate (tabelul 8.4). O apropiere optimă există între cele două serii de valori (cea mai mare abatere: sub 5%).

Valorile estimate sînt valori probabile, grevate de incertitudine, dependența între x și y fiind de natură stohastică.

Dacă funcția de modelare a dependenței statistice „a rezistat" verificării, nu rămîne decît să se extrapoleze pentru perioada de previziune.

Evident, un model de corelație nu poate fi proiectat cu eficiență pe perioade prea îndepărtate, deoarece gradul de influență a factorilor luați în considerare variază odată cu variația factorilor înșiși și nu este exclus să intervină factori și condiții noi, necunoscute la elaborarea modelului.

[34] Cu r s-a notat coeficientul empiric de corelație și cu ρ — coeficientul teoretic de corelație

Scanned with OKEN Scanner

ANEXE

Valori tabelare ale funcțiilor statistice uzuale

Scanned with OKEN Scanner

Anexa 1

REPARTIȚIA POISSON

$$P(\lambda\ ;\ k) = e^{-\lambda} \cdot \frac{\lambda}{k!}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
0	0,904837	0,818731	0,740818	0,670320	0,606529	0,548812	0,496585	0,449329	0,406570	0,367879
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287	0,347610	0,359463	0,365913	0,367879
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786	0,121663	0,143785	0,164661	0,183940
3	0,000151	0,001092	0,003334	0,007150	0,012636	0,019757	0,028388	0,038343	0,049398	0,061313
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964	0,004968	0,007669	0,011115	0,015528
5		0,000002	0,000015	0,000057	0,000158	0,000356	0,000696	0,001227	0,002001	0,003066
6			0,000001	0,000004	0,000013	0,000036	0,000081	0,000164	0,000300	0,000511
7					0,000001	0,000003	0,000008	0,000019	0,000039	0,000073
8							0,000001	0,000002	0,000004	0,000009
										0,000001

Scanned with OKEN Scanner

2	3	4	5	6	7	8	9	10	11
0,135335	0,049787	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045	0,000017
0,270671	0,149361	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111	0,000454	0,000184
0,270671	0,224042	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998	0,002270	0,001010
0,180447	0,224042	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994	0,007567	0,003705
0,090224	0,168031	0,195367	0,175467	0,133853	0,091226	0,057252	0,033737	0,018917	0,010189
0,036089	0,100819	0,156293	0,175457	0,160623	0,127717	0,091604	0,060727	0,037833	0,022415
0,012030	0,050409	0,104196	0,146223	0,160623	0,149003	0,122138	0,091090	0,063005	0,041095
0,003437	0,021604	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116	0,090079	0,064577
0,000859	0,008102	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756	0,112599	0,088794
0,000191	0,002701	0,013231	0,036266	0,068838	0,101405	0,124077	0,131756	0,125110	0,108526
0,000038	0,000810	0,005292	0,018133	0,041303	0,070983	0,099262	0,118580	0,125110	0,119378
0,000007	0,000221	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020	0,113736	0,119378
0,000001	0,000055	0,000642	0,003434	0,011264	0,026350	0,048127	0,072765	0,094780	0,109430
	0,000013	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376	0,072908	0,092595
	0,000003	0,000056	0,000472	0,002228	0,007094	0,016924	0,032384	0,052077	0,072735
	0,000001	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431	0,034718	0,053352
		0,000004	0,000049	0,000334	0,001448	0,004513	0,010930	0,021699	0,036680
		0,000001	0,000014	0,000118	0,000596	0,002124	0,005786	0,012764	0,023734
			0,000004	0,000039	0,000232	0,000944	0,002893	0,007091	0,014504
			0,000001	0,000012	0,000085	0,000397	0,001370	0,003732	0,008397
				0,000004	0,000030	0,000159	0,000617	0,001866	0,004618
				0,000001	0,000010	0,000061	0,000264	0,000889	0,002419
					0,000003	0,000022	0,000108	0,000404	0,001210
					0,000001	0,000008	0,000042	0,000176	0,000578
						0,000003	0,000016	0,000073	0,000265
						0,000001	0,000006	0,000029	0,000117
							0,000002	0,000011	0,000049
							0,000001	0,000004	0,000020
								0,000001	0,000008
								0,000001	0,000003
									0,000001

Scanned with OKEN Scanner

Anexa 2

DENSITATEA DE PROBABILITATE A REPARTIȚIEI NORMALE NORMATE $N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,1814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,4268	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0442	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0094	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Scanned with OKEN Scanner

Anexa 3

FUNCȚIA DE REPARTIȚIE NORMALĂ NORMATĂ (LAPLACE) $N(0,1)$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

X	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7380	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Scanned with OKEN Scanner

Anexa 4

REPARTIȚIA χ^2

n	0,95	0,99	n	0,95	0,99
1	3,84	6,63	51	68,7	77,4
2	5,99	9,21	52	69,8	78,6
3	7,81	11,3	53	71,0	79,8
4	9,49	13,3	54	72,2	81,1
5	11,1	15,1	55	73,3	82,3
6	12,6	16,8	56	74,5	83,5
7	14,1	18,5	57	75,6	84,7
8	15,5	20,1	58	76,8	86,0
9	16,9	21,7	59	77,9	87,2
10	18,3	23,2	60	79,1	88,4
11	19,7	24,7	61	80,2	89,6
12	21,0	26,2	62	81,4	90,8
13	22,4	27,7	63	82,5	92,0
14	23,7	29,1	64	83,7	93,2
15	25,0	30,6	65	84,8	94,4
16	26,3	32,0	66	86,0	95,6
17	27,6	33,4	67	87,1	96,8
18	28,9	34,8	68	88,3	98,0
19	30,1	36,2	69	89,4	99,2
20	31,4	37,6	70	90,5	100,4
21	32,7	38,9	71	91,7	101,6
22	33,9	40,3	72	92,8	102,8
23	35,2	41,6	73	93,9	104,0
24	36,4	43,0	74	95,1	105,2
25	37,7	44,3	75	96,2	106,4
26	38,9	45,6	76	97,4	107,6
27	40,1	47,0	77	98,5	108,8
28	41,3	48,3	78	99,6	110,0
29	42,6	49,6	79	100,7	111,1
30	43,8	50,9	80	101,9	112,3
31	45,0	52,2	81	103,0	113,5
32	46,2	53,5	82	104,1	114,7
33	47,4	54,8	83	105,3	115,9
34	48,6	56,1	84	106,4	117,1
35	49,8	57,3	85	107,5	118,2
36	51,0	58,6	86	108,6	119,4
37	52,2	59,9	87	109,8	120,6
38	53,4	61,2	88	110,9	121,8
39	54,6	62,4	89	112,0	122,9
40	55,8	63,7	90	113,1	124,1
41	56,9	65,0	91	114,3	125,3
42	58,1	66,2	92	115,4	126,5
43	59,3	67,5	93	116,5	127,6
44	60,5	68,7	94	117,6	128,8
45	61,7	70,0	95	118,8	130,0
46	62,8	71,2	96	119,9	131,1
47	64,0	72,4	97	121,0	132,3
48	65,2	73,7	98	122,1	133,5
49	66,3	74,9	99	123,2	134,6
50	67,5	76,2	100	124,3	135,8

Notă. Pentru valorile mari ale lui n putem să ne folosim de expresia $\sqrt{2\chi^2} - \sqrt{2n-1}$ drept abatere normată exprimată în unitățile erorii standard. În acest caz abaterea normată este folosită pentru determinarea ariei delimitate de o parte a curbei.

Scanned with OKEN Scanner

Anexa 5

VALORILE LUI t

n	$P = 0,05$	0,02	0,01
1	12,706	31,821	63,657
2	4,303	6,965	9,925
3	3,182	4,541	5,841
4	2,776	3,747	4,604
5	2,571	3,365	4,032
6	2,447	3,143	3,707
7	2,365	2,998	3,499
8	2,306	2,896	3,355
9	2,262	2,821	3,250
10	2,228	2,764	3,169
11	2,201	2,718	3,106
12	2,179	2,681	3,055
13	2,160	2,650	3,012
14	2,145	2,624	2,977
15	2,131	2,602	2,947
16	2,120	2,583	2,921
17	2,110	2,567	2,898
18	2,101	2,552	2,878
19	2,093	2,539	2,861
20	2,086	2,528	2,845
21	2,080	2,518	2,831
22	2,074	2,508	2,819
23	2,069	2,500	2,807
24	2,064	2,492	2,797
25	2,060	2,485	2,787
26	2,056	2,479	2,779
27	2,052	2,473	2,771
28	2,048	2,467	2,763
29	2,045	2,462	2,756
30	2,042	2,457	2,750
∞	1,95996	2,32634	2,57582

Scanned with OKEN Scanner

Anexa 6

VALORILE RAPORTULUI DISPERSIILOR F CORESPUNZĂTOARE DIFERITELOR

A. **Nivel de semnificaţie 5%**

ν_2 \ ν_1	1	2	3	4	5
1	161,4	199,5	215,7	224,6	230,2
2	18,51	19,00	19,16	19,25	19,30
3	10,13	9,55	9,28	9,12	9,01
4	7,71	6,94	6,59	6,39	6,26
5	6,61	5,79	5,41	5,19	5,05
6	5,99	5,14	4,76	4,53	4,39
7	5,58	4,74	4,35	4,12	3,97
8	5,32	4,46	4,07	3,84	3,69
9	5,12	4,26	3,86	3,63	3,48
10	4,96	4,10	3,71	3,48	3,33
11	4,84	3,98	3,59	3,36	3,20
12	4,75	3,88	3,49	3,26	3,11
13	4,67	3,80	3,41	3,18	3,02
14	4,60	3,74	3,34	3,11	2,96
15	4,54	3,68	3,29	3,06	2,90
16	4,49	3,63	3,24	3,01	2,85
17	4,45	3,59	3,20	2,96	2,81
18	4,41	3,55	3,16	2,93	2,77
19	4,38	3,52	3,13	2,90	2,74
20	4,35	3,49	3,10	2,87	2,71
21	4,32	3,47	3,07	2,84	2,68
22	4,30	3,44	3,05	2,82	2,66
23	4,28	3,42	3,03	2,80	2,64
24	4,26	3,40	3,01	2,78	2,62
25	4,24	3,38	2,99	2,76	2,60
26	4,22	3,37	2,98	2,74	2,59
27	4,21	3,35	2,96	2,73	2,57
28	4,20	3,34	2,95	2,71	2,56
29	4,18	3,33	2,93	2,70	2,54
30	4,17	3,32	2,92	2,69	2,53
40	4,08	3,23	2,84	2,61	2,45
60	4,00	3,15	2,76	2,52	2,37
120	3,92	3,07	2,68	2,45	2,29
∞	3,84	2,99	2,60	2,37	2,21

Întotdeauna ν_1 corespunde celei mai mari dispersii.

Scanned with OKEN Scanner

6	8	12	24	∞
234,0	238,9	243,9	249,0	254,3
19,33	19,37	19,41	19,45	19,50
8,94	8,84	8,74	8,64	8,53
6,16	6,04	5,91	5,77	5,63
4,95	4,82	4,68	4,53	4,36
4,28	4,15	4,00	3,84	3,67
3,87	3,73	3,57	3,41	3,23
3,58	3,44	3,28	3,12	2,93
3,37	3,23	3,07	2,90	2,71
3,22	3,07	2,91	2,74	2,54
3,09	2,95	2,79	2,61	2,40
3,00	2,85	2,69	2,50	2,30
2,92	2,77	2,60	2,42	2,21
2,85	2,70	2,53	2,35	2,13
2,79	2,64	2,48	2,29	2,07
2,74	2,59	2,42	2,24	2,01
2,70	2,55	2,38	2,19	1,96
2,66	2,51	2,34	2,15	1,92
2,63	2,48	2,31	2,11	1,88
2,60	2,45	2,28	2,08	1,84
2,57	2,42	2,25	2,05	1,81
2,55	2,40	2,23	2,03	1,78
2,53	2,38	2,20	2,00	1,76
2,51	2,36	2,18	1,98	1,73
2,49	2,34	2,16	1,96	1,71
2,47	2,32	2,15	1,95	1,69
2,46	2,30	2,13	1,93	1,67
2,44	2,29	2,12	1,91	1,65
2,43	2,28	2,10	1,90	1,64
2,42	2,27	2,09	1,89	1,62
2,34	2,18	2,00	1,79	1,51
2,25	2,10	1,92	1,70	1,39
2,17	2,02	1,83	1,61	1,25
2,09	1,94	1,75	1,52	1,00

Scanned with OKEN Scanner

Anexa 6 (continuare)

VALORILE RAPORTULUI DISPERSIILOR F CORESPUNZĂTOARE DIFERITELOR

***B.* Nivel de semnificaţie 1%**

ν_2 \ ν_1	1	2	3	4	5
1	4052	4999	5403	5625	5764
2	98,49	99,00	99,17	99,25	99,30
3	34,12	30,81	29,46	28,71	28,24
4	21,20	18,00	16,69	15,98	15,52
5	16,26	13,27	12,06	11,39	10,97
6	13,74	10,92	9,78	9,15	8,75
7	12,25	9,55	8,45	7,85	7,46
8	11,26	8,65	7,59	7,01	6,63
9	10,56	8,02	6,99	6,42	6,06
10	10,04	7,56	6,55	5,99	5,64
11	9,65	7,20	6,22	5,67	5,32
12	9,33	6,93	5,95	5,41	5,06
13	9,07	6,70	5,74	5,20	4,86
14	8,86	6,51	5,56	5,03	4,69
15	8,68	6,36	5,42	4,89	4,56
16	8,53	6,32	5,29	4,77	4,44
17	8,40	6,11	5,18	4,67	4,34
18	8,28	6,01	5,09	4,58	4,25
19	8,18	5,93	5,01	4,50	4,17
20	8,10	5,85	4,94	4,43	4,10
21	8,02	5,78	4,87	4,37	4,04
22	7,94	5,72	4,82	4,31	3,99
23	7,88	5,66	4,76	4,26	3,94
24	7,82	5,61	4,72	4,22	3,90
25	7,77	5,57	4,68	4,18	3,86
26	7,72	5,53	4,64	4,14	3,82
27	7,68	5,49	4,60	4,11	3,78
28	7,64	5,45	4,57	4,07	3,75
29	7,60	5,42	4,54	4,04	3,73
30	7,56	5,39	4,51	4,02	3,70
40	7,31	5,18	4,31	3,83	3,51
60	7,08	4,98	4,13	3,65	3,34
120	6,85	4,79	3,95	3,48	3,17
∞	6,64	4,60	3,78	3,32	3,02

Întotdeauna ν_1 corespunde celei mai mari dispersii

Scanned with OKEN Scanner

6	8	12	24	∞
5859	5981	6106	6234	6366
99,33	99,36	99,42	99,46	99,50
27,91	27,49	27,05	26,60	26,12
15,21	14,80	14,37	13,93	13,46
10,67	10,27	9,89	9,47	9,02
8,47	8,10	7,72	7,31	6,88
7,19	6,84	6,47	6,07	5,65
6,37	6,03	5,67	5,28	4,86
5,80	5,47	5,11	4,73	4,31
5,39	5,06	4,71	4,33	3,91
5,07	4,74	4,40	4,02	3,60
4,82	4,50	4,16	3,78	3,36
4,62	4,30	3,96	3,59	3,16
4,46	4,14	3,80	3,43	3,00
4,32	4,00	3,67	3,29	2,87
4,20	3,89	3,55	3,18	2,75
4,10	3,79	3,45	3,08	2,65
4,01	3,71	3,37	3,00	2,57
3,94	3,63	3,30	2,92	2,49
3,87	3,56	3,23	2,86	2,42
3,81	3,51	3,17	2,80	2,36
3,76	3,45	3,12	2,75	2,31
3,71	3,41	3,07	2,70	2,26
3,67	3,36	3,03	2,66	2,21
3,63	3,32	2,99	2,62	2,17
3,59	3,29	2,96	2,58	2,13
3,56	3,26	2,93	2,55	2,10
3,53	3,23	2,90	2,52	2,06
3,50	3,20	2,87	2,49	2,03
3,47	3,17	2,84	2,47	2,01
3,29	2,99	2,66	2,29	1,80
3,12	2,82	2,50	2,12	1,60
2,96	2,66	2,34	1,95	1,38
2,80	2,51	2,18	1,79	1,00

Scanned with OKEN Scanner

Anexa 7

VALORILE LUI z CORESPUNZĂTOARE DIFERITELOR NIVELE DE SEMNIFICAȚIE

A. **Nivel de semnificație 5%**

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	2,5421	2,6479	2,6870	2,7071	2,7194	2,7276	2,7380	2,7484	2,7588	2,7693
2	1,4592	1,4722	1,4765	1,4787	1,4800	1,4808	1,4819	1,4830	1,4840	1,4851
3	1,1577	1,1284	1,1137	1,1051	1,0994	1,0953	1,0899	1,0842	1,0781	1,0716
4	1,0212	0,9690	0,9429	0,9272	0,9168	0,9093	0,8993	0,8885	0,8767	0,8639
5	0,9441	0,8777	0,8441	0,8236	0,8097	0,7997	0,7862	0,7714	0,7550	0,7368
6	0,8948	0,8188	0,7798	0,7558	0,7394	0,7274	0,7112	0,6931	0,6729	0,6499
7	0,8606	0,7777	0,7347	0,7080	0,6896	0,6761	0,6576	0,6369	0,6134	0,5862
8	0,8355	0,7475	0,7014	0,6725	0,6525	0,6378	0,6175	0,5945	0,5682	0,5371
9	0,8163	0,7242	0,6757	0,6450	0,6238	0,6080	0,5862	0,5613	0,5324	0,4979
10	0,8012	0,7058	0,6553	0,6232	0,6009	0,5843	0,5611	0,5346	0,5035	0,4657
11	0,7889	0,6909	0,6387	0,6055	0,5822	0,5648	0,5406	0,5126	0,4795	0,4387
12	0,7788	0,6786	0,6250	0,5907	0,5666	0,5487	0,5234	0,4941	0,4592	0,4156
13	0,7703	0,6682	0,6134	0,5783	0,5535	0,5350	0,5089	0,4785	0,4419	0,3957
14	0,7630	0,6594	0,6336	0,5677	0,5423	0,5233	0,4964	0,4649	0,4269	0,3782
15	0,7568	0,6518	0,5950	0,5585	0,5326	0,5131	0,4855	0,4532	0,4138	0,3628
16	0,7514	0,6451	0,5876	0,5505	0,5241	0,5042	0,4760	0,4428	0,4022	0,3490
17	0,7466	0,6393	0,5811	0,5434	0,5166	0,4964	0,4676	0,4337	0,3919	0,3366
18	0,7424	0,6341	0,5753	0,5371	0,5099	0,4894	0,4602	0,4255	0,3827	0,3253
19	0,7386	0,6295	0,5701	0,5315	0,5040	0,4832	0,4535	0,4182	0,3743	0,3151
20	0,7352	0,6254	0,5654	0,5265	0,4986	0,4776	0,4474	0,4116	0,3668	0,3057
21	0,7322	0,6216	0,5612	0,5219	0,4938	0,4725	0,4420	0,4055	0,3599	0,2971
22	0,7294	0,6182	0,5574	0,5178	0,4894	0,4679	0,4370	0,4001	0,3536	0,2892
23	0,7269	0,6151	0,5540	0,5140	0,4854	0,4636	0,4325	0,3950	0,3478	0,2818
24	0,7246	0,6123	0,5508	0,5106	0,4817	0,4598	0,4283	0,3904	0,3425	0,2749
25	0,7225	0,6097	0,5478	0,5074	0,4783	0,4562	0,4244	0,3862	0,3376	0,2685
26	0,7205	0,6073	0,5451	0,5045	0,4752	0,4529	0,4209	0,3823	0,3330	0,2625
27	0,7187	0,6151	0,5427	0,5017	0,4723	0,4499	0,4176	0,3786	0,3287	0,2569
28	0,7171	0,6030	0,5403	0,4992	0,4696	0,4471	0,4146	0,3752	0,3248	0,2516
29	0,7155	0,6011	0,5382	0,4969	0,4671	0,4444	0,4117	0,3720	0,3211	0,2466
30	0,7141	0,5994	0,5362	0,4947	0,4648	0,4420	0,4090	0,3691	0,3176	0,2419
60	0,6933	0,5738	0,5073	0,4632	0,4311	0,4064	0,3702	0,3255	0,2654	0,1644
∞	0,6729	0,5486	0,4787	0,4319	0,3974	0,3706	0,3309	0,2804	0,2085	0,0000

Scanned with OKEN Scanner

Anexa 7 (continuare)

VALORILE LUI z CORESPUNZĂTOARE DIFERITELOR NIVELE DE SEMNIFICAȚIE

B. **Nivel de semnificație 1%**

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	4,1535	4,2585	4,2974	4,3175	4,3297	4,3379	4,3482	4,3585	4,3689	4,3794
2	2,2950	2,2976	2,2984	2,2988	2,2991	2,2992	2,2994	2,2997	2,2999	2,3001
3	1,7649	1,7140	1,6915	1,6786	1,6703	1,6645	1,6569	1,6489	1,6404	1,6314
4	1,5270	0,4452	1,4075	1,3856	1,3711	1,3609	1,3473	1,3327	1,3170	1,3000
5	1,3943	1,2929	1,2449	1,2164	1,1974	1,1838	1,1656	1,1457	1,1239	1,0997
6	1,3103	1,1955	1,1401	1,1068	1,0843	1,0680	1,0460	1,0218	0,9948	0,9643
7	1,2526	1,1281	1,0672	1,0300	1,0048	0,9864	0,9614	0,9335	0,9020	0,8658
8	1,2106	1,0787	1,0135	0,9734	0,9459	0,9259	0,8983	0,8673	0,8319	0,7904
9	1,1786	1,0411	0,9724	0,9299	0,9006	0,8791	0,8494	0,8157	0,7769	0,7305
10	1,1535	1,0114	0,9399	0,8954	0,8646	0,8419	0,8104	0,7744	0,7324	0,6816
11	1,1333	0,9874	0,9136	0,8674	0,8354	0,8116	0,7785	0,7405	0,6958	0,6408
12	1,1166	0,9677	0,8919	0,8443	0,8111	0,7864	0,7520	0,7122	0,6649	0,6061
13	1,1027	0,9511	0,8737	0,8248	0,7907	0,7652	0,7295	0,6882	0,6386	0,5761
14	1,0909	0,9370	0,8581	0,8082	0,7732	0,7471	0,7103	0,6675	0,6159	0,5500
15	1,0807	0,9249	0,8448	0,7939	0,7582	0,7314	0,6937	0,6496	0,5961	0,5269
16	1,0719	0,9144	0,8331	0,7814	0,7450	0,7177	0,6791	0,6339	0,5786	0,5064
17	1,0641	0,9051	0,8229	0,7705	0,7335	0,7057	0,6663	0,6199	0,5630	0,4879
18	1,0572	0,8970	0,8138	0,7607	0,7232	0,6950	0,6549	0,6075	0,5491	0,4712
19	1,0511	0,8897	0,8057	0,7521	0,7140	0,6854	0,6447	0,5964	0,5366	0,4560
20	1,0457	0,8831	0,7985	0,7443	0,7058	0,6768	0,6355	0,5864	0,5253	0,4421
21	1,0408	0,8772	0,7920	0,7372	0,6984	0,6690	0,6272	0,5773	0,5150	0,4294
22	1,0363	0,8719	0,7860	0,7309	0,6916	0,6620	0,6196	0,5691	0,5056	0,4176
23	1,0322	0,8670	0,7806	0,7251	0,6855	0,6555	0,6127	0,5615	0,4969	0,4068
24	1,0285	0,8626	0,7757	0,7197	0,6799	0,6496	0,6064	0,5545	0,4890	0,3967
25	1,0251	0,8585	0,7712	0,7148	0,6747	0,6442	0,6006	0,5481	0,4816	0,3872
26	1,0220	0,8548	0,7670	0,7103	0,6699	0,6392	0,5952	0,5422	0,4748	0,3784
27	1,0191	0,8513	0,7631	0,7062	0,6655	0,6346	0,5902	0,5367	0,4685	0,3701
28	1,0164	0,8481	0,7595	0,7023	0,6614	0,6303	0,5856	0,5316	0,4626	0,3624
29	1,0139	0,8451	0,7562	0,6987	0,6576	0,6263	0,5813	0,5269	0,4570	0,3550
30	1,0116	0,8423	0,7531	0,6954	0,6540	0,6226	0,5773	0,5224	0,4519	0,3481
60	0,9784	0,8025	0,7086	0,6472	0,6028	0,5687	0,5189	0,4574	0,3746	0,2352
∞	0,9452	0,7636	0,6651	0,5999	0,5525	0,5152	0,4604	0,3908	0,2913	0,0000

Scanned with OKEN Scanner

Anexa 8

VALORILE COEFICIENTULUI DE CORELAȚIE PENTRU DIFERITE NIVELE DE SEMNIFICAȚIE

n	$P = 0,05$	0,02	0,01
1	0,996917	0,9995066	0,9998766
2	0,95000	0,98000	0,990000
3	0,8783	0,93433	0,95873
4	0,8114	0,8822	0,91720
5	0,7545	0,8329	0,8745
6	0,7067	0,7887	0,8343
7	0,6664	0,7498	0,7977
8	0,6319	0,7155	0,7646
9	0,6021	0,6851	0,7348
10	0,5760	0,6581	0,7079
11	0,5529	0,6339	0,6835
12	0,5324	0,6120	0,6614
13	0,5139	0,5923	0,6411
14	0,4973	0,5742	0,6226
15	0,4821	0,5577	0,6055
16	0,4683	0,5425	0,5897
17	0,4555	0,5285	0,5751
18	0,4438	0,5155	0,5614
19	0,4329	0,5034	0,5487
20	0,4227	0,4921	0,5368
25	0,3809	0,4451	0,4869
30	0,3494	0,4093	0,4487
35	0,3246	0,3810	0,4182
40	0,3044	0,3578	0,3932
45	0,2875	0,3384	0,3721
50	0,2732	0,3218	0,3541
60	0,2500	0,2948	0,3248
70	0,2319	0,2737	0,3017
80	0,2172	0,2565	0,2830
90	0,2050	0,2422	0,2673
100	0,1946	0,2301	0,2540

Scanned with OKEN Scanner

Anexa 9

RELAȚIA ÎNTRE r ȘI Z' PENTRU VALORILE LUI z' DE LA 0 LA 3

z'	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898
0,1	0,0997	0,1096	0,1194	0,1293	0,1391	0,1489	0,1587	0,1684	0,1781	0,1878
0,2	0,1974	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821
0,3	0,2913	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714
0,4	0,3800	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542
0,5	0,4621	0,4700	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299
0,6	0,5370	0,5441	0,5511	0,4581	0,5649	0,5717	0,5748	0,5850	0,5915	0,5980
0,7	0,6044	0,6107	0,6169	0,6231	0,6291	0,6352	0,6411	0,6469	0,6527	0,6584
0,8	0,6640	0,6696	0,6751	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114
0,9	0,7163	0,7211	0,7259	0,7306	0,7352	0,7398	0,7443	0,7487	0,7531	0,7574
1,0	0,7616	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969
1,1	0,8005	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306
1,2	0,8337	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591
1,3	0,8617	0,8643	0,8668	0,8693	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832
1,4	0,8854	0,8875	0,8896	0,8917	0,8937	0,8957	0,8977	0,8996	0,9015	0,9033
1,5	0,9052	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9202
1,6	0,9217	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9342
1,7	0,9354	0,9367	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458
1,8	0,9468	0,9478	0,9498	0,9488	0,9508	0,9518	0,9527	0,9536	0,9545	0,9545
1,9	0,9562	0,9571	0,9579	0,9587	0,9595	0,9603	0,9611	0,9619	0,9626	0,9633
2,0	0,9640	0,9647	0,9654	0,9661	0,9668	0,9674	0,9680	0,9687	0,9693	0,9699
2,1	0,9705	0,9710	0,9716	0,9722	0,9727	0,9732	0,9738	0,9743	0,9748	0,9753
2,2	0,9757	0,9762	0,9767	0,9771	0,9776	0,9780	0,9785	0,9789	0,9793	0,9797
2,3	0,9801	0,9805	0,9809	0,9812	0,9816	0,9820	0,9823	0,9827	0,9830	0,9834
2,4	0,9837	0,9840	0,9843	0,9846	0,9849	0,9852	0,9855	0,9858	0,9861	0,9863
2,5	0,9866	0,9869	0,9871	0,9874	0,9876	0,9879	0,9881	0,9884	0,9886	0,9888
2,6	0,9890	0,9892	0,9895	0,9897	0,9899	0,9901	0,9903	0,9905	0,9906	0,9908
2,7	0,9910	0,9912	0,9914	0,9915	0,9917	0,9919	0,9920	0,9922	0,9923	0,9925
2,8	0,9926	0,9928	0,9929	0,9931	0,9932	0,9933	0,9935	0,9936	0,9937	0,9938
2,9	0,9940	0,9941	0,9942	0,9943	0,9944	0,9945	0,9946	0,9947	0,9949	0,9950

Scanned with OKEN Scanner

Anexa 10

NUMERE ÎNTÎMPLĂTOARE

	5	10	15	20	25	30	35	40	45	50
	13 407	62 899	78 937	90 525	25 033	56 358	78 902	47 008	72 488	57 949
	50 230	63 237	94 083	93 634	71 652	02 656	57 532	60 307	91 619	48 916
	84 980	62 458	09 703	78 397	66 179	46 982	67 619	39 254	90 763	74 056
	22 116	33 646	17 545	31 321	65 772	86 506	09 811	82 848	92 211	51 178
5	68 645	15 068	56 898	87 021	40 115	27 524	42 221	88 293	67 592	06 430
	26 518	39 122	96 561	56 004	50 260	68 648	85 596	83 979	09 041	62 350
	36 493	41 666	27 871	71 329	69 212	57 932	65 281	57 233	07 732	58 439
	77 402	12 994	59 892	85 581	70 823	53 338	34 405	67 080	16 568	00 854
	83 679	97 154	40 341	84 741	08 967	73 287	94 952	59 008	95 774	44 927
10	71 802	39 356	02 981	89 107	79 788	51 330	37 129	31 898	34 011	43 304
	57 494	72 484	22 676	44 311	15 356	05 348	03 582	66 183	68 392	86 844
	73 364	38 416	93 128	10 297	11 419	82 937	84 389	88 273	96 010	09 843
	14 499	83 965	75 403	18 002	45 068	54 257	18 085	92 625	60 911	39 137
	40 747	03 084	07 734	88 940	88 722	85 717	73 810	79 866	84 853	68 647
15	42 237	59 122	92 855	62 097	81 276	06 318	81 607	00 565	56 626	77 422
	32 934	60 227	58 707	44 858	36 081	79 981	01 291	68 707	45 427	82 145
	05 764	14 284	73 069	80 830	17 231	42 936	48 472	18 782	51 646	37 564
	32 706	94 879	93 188	66 049	25 988	46 656	35 365	13 800	83 745	40 141
	22 190	27 559	95 668	53 261	21 676	98 943	43 618	42 110	93 402	93 997
20	81 616	15 641	94 921	95 970	63 506	22 007	29 966	38 144	62 556	07 864
	26 099	65 801	69 870	84 446	58 248	21 282	56 938	54 729	67 757	68 412
	71 874	61 692	80 001	21 430	02 305	59 741	34 262	15 157	27 545	14 522
	08 774	29 689	42 245	51 903	69 179	96 682	91 819	60 812	47 631	50 609
	37 294	92 028	56 850	83 380	05 912	29 830	37 612	15 593	73 198	99 287
25	33 912	37 996	78 967	57 201	66 916	73 998	54 289	07 147	84 313	51 938
	63 610	61 475	26 980	23 804	54 972	72 068	19 403	53 756	04 281	98 022
	01 570	41 701	30 282	54 647	06 077	29 354	95 704	75 928	21 811	88 274
	24 159	77 787	38 973	82 178	46 802	90 245	01 805	23 906	96 559	06 785
	92 834	52 941	88 301	22 127	23 459	40 229	74 678	21 859	98 645	72 388
30	16 178	60 063	59 284	16 279	48 003	44 634	08 623	32 752	40 472	05 470
	81 808	32 980	80 660	98 391	62 243	19 678	39 551	18 398	36 918	43 543
	28 628	82 072	04 854	52 809	86 608	68 017	11 120	28 638	72 850	03 650
	62 249	65 757	12 273	91 261	96 983	15 082	83 851	77 682	81 728	52 157
	84 541	99 891	01 585	96 711	29 712	02 877	70 955	59 693	26 838	96 011
35	89 052	39 061	99 811	69 831	47 234	93 263	47 386	17 462	18 874	74 210
	43 645	89 232	00 384	10 858	21 789	14 093	06 268	46 460	97 660	23 490
	61 618	19 275	40 744	22 482	12 424	98 601	19 089	53 166	41 836	28 205
	68 136	06 546	04 029	47 946	19 526	27 122	42 515	55 048	23 912	81 105
	74 005	34 558	93 779	96 120	01 695	47 720	88 646	73 520	40 050	90 553
40	54 437	88 825	07 943	81 795	31 709	13 358	04 626	64 838	92 133	44 221
	01 990	94 762	89 926	84 764	19 159	95 355	98 213	17 704	47 400	30 837
	02 404	42 408	67 981	43 684	55 467	47 030	42 545	43 920	11 199	36 521
	59 253	71 535	26 149	35 629	87 127	45 581	00 185	01 041	46 662	98 897
	20 471	13 914	99 330	37 938	69 649	57 964	47 149	41 628	78 664	80 727
45	65 946	60 766	74 084	22 484	49 014	89 820	41 310	19 722	07 045	28 808
	00 939	47 818	75 949	44 707	49 105	06 777	31 998	79 942	98 351	10 265
	49 952	29 123	45 950	67 578	13 524	03 023	18 046	75 287	74 989	58 152
	17 328	70 732	46 319	26 950	19 037	02 831	36 558	82 712	05 590	64 941
	19 420	70 215	90 476	76 400	51 553	12 158	14 668	15 656	37 895	94 559
50	19 121	41 190	49 145	05 373	00 755	17 817	22 757	76 116	76 977	94 570

Scanned with OKEN Scanner

Anexa 10 (continuare)

	5	10	15	20	25	30	35	40	45	50
	44 300	56 179	71 202	49 238	83 682	21 289	63 268	74 644	53 625	10 791
	99 403	96 757	34 512	06 475	89 028	00 290	93 766	70 812	98 331	09 611
	78 578	51 589	83 195	56 332	75 076	58 202	58 038	38 817	63 835	13 486
	89 830	60 177	94 550	10 119	09 083	33 398	29 974	67 721	75 037	70 444
55	89 502	83 947	99 940	60 969	79 452	91 472	12 611	41 681	95 285	44 153
	11 187	95 098	50 369	94 874	19 853	06 933	69 767	88 842	35 676	49 766
	47 886	49 549	64 465	14 508	28 215	47 766	03 076	25 940	47 239	93 425
	21 325	89 726	96 964	66 106	68 517	67 954	16 570	72 433	91 514	79 333
	59 927	79 213	96 072	64 540	59 002	26 619	02 930	83 677	26 442	97 346
60	44 232	30 754	59 691	34 893	92 531	70 313	24 969	14 458	91 409	79 369
	15 956	31 379	21 224	20 366	74 348	66 239	32 704	41 018	31 937	84 761
	58 597	14 598	23 589	50 700	96 194	15 831	08 968	45 321	04 207	34 438
	99 185	70 628	95 475	94 156	39 588	57 825	36 521	85 188	64 339	27 460
	20 986	57 081	53 928	47 768	18 313	82 950	12 335	32 298	08 662	54 552
65	75 371	04 678	96 443	72 965	68 012	52 485	55 139	73 430	74 306	85 960
	75 775	60 178	51 110	30 735	29 761	39 565	45 332	13 671	69 405	11 186
	91 592	54 102	25 242	00 063	42 467	23 339	55 311	81 275	08 602	03 508
	16 106	87 812	92 476	07 849	65 510	77 763	33 684	77 092	32 490	40 345
	30 764	57 054	12 611	21 455	01 332	33 101	64 795	56 555	84 390	12 982
70	63 826	14 146	40 993	93 849	49 799	41 080	48 621	29 555	83 633	07 742
	18 068	54 879	36 271	24 773	63 615	60 309	30 550	16 184	71 605	64 267
	32 512	08 155	27 597	59 844	95 648	71 896	63 075	86 078	61 746	26 669
	21 339	99 336	18 200	57 564	39 356	15 173	53 051	87 654	13 346	62 350
	67 124	38 202	10 706	65 647	19 870	15 499	46 540	44 897	53 548	08 521
75	07 444	58 069	35 758	41 422	82 031	35 465	72 508	77 079	81 768	32 854
	78 097	95 307	65 668	21 280	75 514	68 955	57 328	74 675	67 958	37 864
	13 240	79 748	67 309	46 843	19 734	45 248	20 434	77 530	06 735	53 622
	05 412	00 586	33 144	36 553	57 446	66 156	31 637	15 924	71 923	73 089
	54 401	85 120	18 976	42 639	67 159	86 473	79 129	02 003	08 708	65 578
80	61 746	35 493	36 645	23 427	12 223	67 361	19 073	39 770	13 548	64 994
	40 768	91 507	03 755	82 940	64 903	93 497	93 398	76 665	24 527	02 248
	50 483	68 401	34 226	66 607	36 112	14 443	78 651	09 970	86 482	86 338
	55 872	75 721	68 194	70 065	62 636	76 536	61 122	18 934	57 260	66 447
	36 400	13 236	25 341	74 477	31 365	84 502	34 022	78 158	44 113	20 549
85	03 454	13 967	95 650	82 123	91 422	69 308	19 610	20 522	10 510	78 179
	09 501	62 273	27 839	40 721	97 029	28 581	80 024	83 898	12 993	49 612
	68 329	47 970	00 605	23 532	03 540	51 239	37 749	60 779	53 907	46 907
	52 572	88 006	94 321	55 784	59 721	63 064	98 730	84 514	17 055	73 888
	91 241	13 322	35 342	85 566	90 832	53 907	34 692	72 878	24 089	50 360
90	01 993	48 316	20 995	90 202	65 406	96 207	11 940	55 429	31 512	25 399
	08 715	00 444	79 663	62 540	87 229	68 367	73 803	21 153	06 316	90 998
	26 428	32 887	17 086	31 548	24 240	43 789	49 953	01 486	73 662	83 201
	68 335	94 334	26 895	89 630	75 309	07 033	95 049	20 406	53 360	50 756
	73 548	57 010	04 191	94 160	94 533	11 264	05 042	07 881	32 239	16 572
95	57 913	03 318	98 663	25 295	47 813	46 003	85 625	98 613	15 959	04 108
	78 160	94 662	22 608	73 015	23 114	82 185	04 236	70 358	74 165	82 546
	14 433	13 576	42 644	17 019	90 483	80 265	98 711	40 182	77 335	46 841
	58 639	46 140	85 366	74 364	13 940	55 260	35 797	47 739	88 447	14 399
	19 451	65 743	30 209	65 322	91 783	25 563	33 211	93 780	10 690	89 063
100	47 372	98 813	15 017	39 536	97 383	43 262	95 696	31 751	01 724	86 899

Scanned with OKEN Scanner

Bibliografie

[1] NICOLAE CEAUȘESCU, *Raport la cel de-al XI-lea Congres al Partidului Comunist Român*, Editura politică, București, 1974.

[2] *Programul Partidului Comunist Român de făurire a societății socialiste multilateral dezvoltate și înaintare a României spre comunism*, Editura politică, București, 1975.

[3] *Hotărîrea Comitetului Central al Partidului Comunist Român cu privire la perfecționarea sistemului informațional economico-social, introducerea sistemelor de conducere cu mijloace de prelucrare automată a datelor și dotarea economiei naționale cu tehnică de calcul în perioada 1971—1980*, în „Scînteia", 22. IV. 1972.

[4] ACHENWALL, G., *Geschichte der heutigen vornehmsten Europäischen Staaten in Grundrisse*, 3-te verbesserte Auflage, Göttingen im Verlag der Witwe Vandenbrock, 1964.

[5] ALLEN, R.D.G., *Statistics for economists*, Hutschinson University Library, London, 1963.

[6] ALLEN, R.D.G., EDWARD ELY, J., *International Trade Statistics*, Editura John Wiley, New York, 1953.

[7] ALTUNIN, M. P., *Prakticeskoe posobie po obșcei i seliskohoziaistvenoi statistike*, Moskva, 1947.

[8] ARNOLD, R. R., HILL, H.C., NICHOLS, A., V. *Inițiere în prelucrarea datelor*, Editura tehnică, București, 1969.

[9] BAILEY, N. T. J., *Statistical Methods in biology*, English University Press L.T.D., 1959.

[10] BĂRBAT, AL., *Statistica în opera lui George Barițiu*, în vol. *Din istoria statisticii românești*, Direcția Centrală de Statistică, București, 1969.

[11] BĂRBAT, AL., *Teoria statisticii sociale*, Editura didactică și pedagogică, București, 1972.

[12] *Bazele teoriei generale a statisticii (Metode de calcul statistic și analiză economică)*, sub red. prof. Manea Mănescu, Institutul Politehnic, Institutul de Științe Economice, București, 1962.

[13] BERNAL, J. D., *Știința în istoria societății*, Editura politică, București, 1964.

[14] BIJI E., BĂJAN, A., BĂDIȚĂ, M., *Culegere de probleme de teoria statisticii*, Academia de Studii Economice, București, 1973.

[15] BIJI, M., BIJI, E., *Curs de teoria statisticii*, Academia de Studii Economice, București, 1971.

[16] BIJI, M., *Ion Ionescu de la Brad un statistician progresist român*, în „Revista de statistică" nr. 5/1955.

[17] BLOCK, M., *Traité théorique et pratique de statistique*, Paris, 1878.

Scanned with OKEN Scanner

[18] BOWLEY, A. L., *Elements de statistique*, Paris 1929.
[19] BRIMO, A., *Les méthodes des sciences sociales*, Editions Montchrestien, Paris, 1972.
[20] BULGARU, M., *Dimitrie Cantemir reprezentant de seamă al statisticii descriptive*, în „Revista de statistică" nr. 11/1973.
[21] CAPĂTĂ, M., *Balanța statistică a legăturilor dintre ramuri. Probleme metodologice*, Editura politică, București, 1974.
[22] CAPĂTĂ, M., *Despre rolul și locul statisticii în cadrul sistemului informațional economico-social*, „Revista de statistică" nr. 11/1973.
[23] CAPĂTĂ, M., LECHKUN, I., MARIN, M., TEODOROF, I., DINCULESCU, ȘT., FILIPIDESCU, AL. *Perfecționarea sistemului informațional statistic. Proiect-cadru*, Direcția Centrală de Statistică, București, 1975.
[24] CAPĂTĂ, M., LECHKUN, I., MARIN, M., TEODOROF, I., DINCULESCU, ȘT., FILIPIDESCU, AL., *Studii privind principiile, cerințele și etapele îmbunătățirii sistemului de indicatori ai dezvoltării economico-sociale*, Direcția Centrală de Statistică, București, aprilie 1975.
[25] CERNEA, E., *Alexandru Papiu-Ilarian*, în *Din gîndirea politico-juridică în România*, Editura științifică, București, 1974.
[26] CIHOSKI, ȘT., *Statistica generală (Metodologia statistică)*, Atelierele graf. Socec & Co., București, 1910.
[27] *Colectiv de autori, Studii de statistică informațională.* Comunicări și referate prezentate la Seminarul de aplicații a metodelor statistico-matematice în economie, Direcția Centrală de Statistică, București, 1975.
[28] CRISTUREANU, T., IVĂNESCU, I., NEGURĂ, I., STOIANOVICI, G., *Sistemul informațional și calculul statistico-economic în comerțul exterior*, Editura științifică, București, 1974.
[29] CROXTON, F .E., COWDEN, D. J., KLEIN, S., *Applied General Statistics*, ed. a 3-a, Prentice-Hall, New York, 1967.
[30] *Curs de prognoză*, Laboratorul de cercetări prospective al Universității București. C.I.D.S.P.
[31] DARMOIS, G., *Statistique et aplications*, ed. a 5-a, Paris, 1957.
[32] DEMETRESCU, M. C., *Elasticitatea cererii populației*, Editura Academiei R.S.R., București, 1967.
[33] DEMETRESCU, R., *Un curs de statistică din anul 1856*, în „Revista de statistică" nr. 1/1970.
[34] *Dicționar de economie politică.* Editura politică, București, 1974.
[35] *Dicționar enciclopedic român*, Editura politică, București, 1964.
[36] *Dicționar statistic-economic*, ediția a 2-a, Direcția Centrală de Statistică, București, 1969.
[37] DOBRESCU, EM., *Ritmul creșterii economice*, Editura politică, București, 1968.
[38] DRUJININ, N. K., *Matematiceskaia statistica v ekonomike*, Editura „Statistica", Moskva, 1971.
[39] DUBOS, J., *Méthodologie économique et technique statistique*, t. I — Statistique, Montpellier, 1965.
[40] DUBOS, J., *Méthodologie économique et téchnique statistique*, t. 4, Les problémes de liaisons entre les variables, I.P.A., Montpellier, 1965.
[41] DUFAU, A. P., *Traité de statistique ou théorie de l'étude des lois d'après lesquelles se développe les faits sociaux*, Paris, 1840.
[42] DUNIN-BARKOVSKI, I. V., SMIRNOV, I. V., *Teoria verotatnostei i matematiceskaia statistica v tehnike (obșciaea ciasti)*, Moskva, 1955.
[43] ENGELS, FR., *Scrisoare către K. Schmidt*, în *K. Marx-Fr. Engels, Scrisori alese.*

Scanned with OKEN Scanner

[44] EZEKIEL, M., FOX, A., *Methods of Correlation and Regression Analysis*, ed. a 3-a, New York, 1959.
[45] FOULQUIÉ, P., SAINT-JEAN, R., *Dictionaire de la langue philosophique*, Paris, 1969.
[46] GALVANI, L., *Statistica*, în *Enciclopedia italiana di scienze, lettere ed arti*, Istituto della enciclopedia it. fondata de Giovanni Treccane, Roma, 1950.
[47] GEORGESCU-ROEGEN, N. ST., *Metoda statistică. Elemente de statistică matematică*, M. O., Imprimeria națională, București, 1933.
[48] GIOJA, M., *Filosofia della statistica*, Mendrisio, 1839.
[49] GIRAULT, M., *Éléments de méthodologie statistique*, Dunod, Paris, 1967.
[50] GUERBER, L., *Statistique descriptive*, Précis Dalloz, Paris, 1971.
[51] GUITTON, H., *Statistique*, Dalloz, Paris, 1967.
[52] HALBWACHS, M., *La théorie de l'homme moyen. Essai sur Quételet et la statistique moral* Libr. Felix Alcan, Paris, 1943.
[53] HAȘIGAN, D., MARINESCU, I., *Grafice și elemente de calcul grafic*, Editura științifică, București, 1968.
[54] HEUSCHLING, X., *Manuel de statistique ethnographique universelle, précedé d'une introduction théorique d'après l'état actuel de la science*, Bruxelles, 1847.
[55] HORVATH, A. R., *Le concept de statistique internationale et son évolution historique*, în „International Statistical Review" nr. 3/1972.
[56] HUBER, M., *Cours de statistique appliquée aux affaires*, II, Paris, 1943.
[57] IASTREMSKI, B. S., *Nekotorîie voprosî matematiceskoi statistiki*, Gosstatizdat, Moskva, 1961.
[58] ILIESCU, I., *„Neinversarea" funcțiilor de regresie și limitele previzionale ale coeficienților de elasticitate*, în „Revista de statistică" nr. 5/1972.
[59] IONESCU DE LA BRAD, I., *Povățuiri pentru catagrafia Moldovei precedate de oarecare elemente de statistică și publicate de Direcția Centrală de Statistică din Ministerul Interior al Moldovei*, Tip. „Buciumul Român", Iași, 1859.
[60] IOSIFESCU, M., MIHOC, GH., THEODORESCU, R., *Teoria probabilităților și statistica matematică*, Editura tehnică București, 1966.
[61] JULIN, AR., *Principes de statistique théorique et appliquée*, Bruxelles, 1921.
[62] KENDALL, M. G., BUKLAND, W. R., *A dictionary of Statistical Terms*, Oliver & Boyd, Londra, 1966.
[63] KOLMOGOROV, A. N., *Gründbegriffe der Warscheinlichkeitsrechnung*, în „Mathematische Ergebnisse" nr. 3, Berlin 1933.
[64] LAPLACE, P. S., *Essai philosophique sur les probabilités*, 7-e éd. Bruxelles, 1840.
[65] *Lecții de statistică matematică*, Direcția Centrală de Statistică, București, 1964.
[66] LENIN, V. I., *Opere complete*, vol. 26, Editura politică, București, 1964.
[67] *Lexikon der Wirtschaft. Rechnungsführung und Statistik*, Verlag „Die Wirtschaft", Berlin-R. D. Germană, 1974.
[68] LOVIN, M., PREOTEASA, P., POPOVICI, P. C., *Informatică*, Editura didactică și pedagogică, București, 1973.
[69] LUKOMSKI, I. I., *Teoria korrelații i elo primenenie k analizu proizvodstva*, Gosstatizdat, Moskva, 1961.
[70] MĂNESCU, M. și colectiv, *Statistica nivelului de trai al populației. Tehnici de calcul statistic și analiză economică*, Editura didactică și pedagogică, București, 1967.
[71] MĂNESCU, M., *Statistica și sociologia*, în *Teorie și metodă în științele sociale*, Editura politică, București, 1965.
[72] MARCH, L., *Les principes de la méthode statistique*, Paris, 1930.
[73] MARINESCU, I., *Considerații teoretice cu privire la metodele de calcul al ritmului mediu de dezvoltare*, „Revista de statistică" nr. 7/1960.

Scanned with OKEN Scanner

74] MARINESCU, I., MOINEAGU, C., NICULESCU, R., RANCU, N., URSEANU, V., *Elemente de statistică matematică și aplicațiile ei*, Editura științifică, București, 1966.
[75] MARȚIAN, D. POP, *Articol program*, în „Anale statistice" nr. 1/1860.
[76] MARȚIAN, D. POP, *Opere economice*, ed. îngrijită de N. Marcu și Z. Ornea, Editura științifică, București, 1961.
[77] MARȚIAN, D. POP, *Statistica justițială*, în „Anale statistice" nr. 17—20/1864.
[78] MARX, K., *Capitalul*, vol. 1, Editura politică, București 1960; vol. 3, p. II, E.S.P.L., București, 1955.
[79] MARX, K., *Contribuții la critica economiei politice*, Editura politică, București, 1969.
[80] MENGES, G., *Grundriss der Statistik*. Teil 1: Theorie, Westdeutscher Verlag, Köln und Opladen, 1968; Teil 2: Daten (în colab. cu H. J. Skala), aceeași editură, 1973.
[81] *Metode statistice de cercetare a corelațiilor în economie*, Editura Academiei R.S.R., București, 1967.
[82] MIHOC, GH., *Elemente de calculul probabilităților*, Editura tehnică, București, 1954.
[83] MIHOC, GH., IOSIFESCU, M., URSEANU, V., *Elemente de teoria probabilităților și aplicațiile ei*, Editura științifică, București, 1966.
[84] MIHOC, GH., URSEANU, V., *Concepte și principii de bază ale teoriei probabilităților*, în „Revista de statistică" nr. 2—7, 10/1971.
[85] MIHOC, GH., URSEANU, V., *Matematici aplicate în statistică*, Editura Academiei R.S.R., București, 1962.
[86] MIHOC, GH., URSEANU, V., *Teoria sondajului*, în „Revista de statistică" nr. 3—11/1972; 1—3/1973.
[87] MILLS, FR. C., *Metode statistice*, ediția a 3-a, Columbia University, New York, 1955, traducere, Direcția Centrală de Statistică, București.
[88] MOINEAGU, C., *Curs de statistică judiciară*, Centrul de multiplicare al Universității din București, 1970.
[89] MOINEAGU, C., *Metoda corelației*, în *Metode statistice de cercetare a corelațiilor în economie*, Editura Academiei R.S.R., București, 1967.
[90] MOINEAGU, C., *Modelarea corelațiilor în economie*, Editura științifică, București, 1974.
[91] MOINEAGU, C., BELU, GH., *Aplicarea metodei corelației statistice la descompunerea influențelor pe factori*, în *Studii de statistică*, Direcția Centrală de Statistică, București, 1964.
[92] MOINEAGU, C., GHEȚIE, C., POPESCU, O., *Aplicații ale metodei corelației statistice la analiza factorilor prețului de cost al producției de lapte*, în „Studii și cercetări de calcul economic și cibernetică economică" nr. 2/1967.
[93] MOREAU DE JONNÈS, A., *Eléments de statistique*, 2-e éd., Paris, 1856.
[94] MORICE, E., *Dictionnaire de statistique*, Dunod, Paris, 1968.
[95] MORICE, E., CHARTIER, F., *Méthode statistique*, Ière partie — Elaboration des statistiques, IIème partie — Analyse statistique, Imprimerie Nationale, Paris, 1954.
[96] NĂDEJDE ILEANA, NEGURĂ, I., *Aplicații ale metodei indicilor în comerțul exterior*, în „Revista de statistică" nr. 1/1962.
[97] NEGURĂ, I., *Probleme ale comparațiilor statistice pe plan internațional*, în *Studii de statistică*, Direcția Centrală de Statistică, 1969.
[98] NEGURĂ, I., *Procedee de ajustare a seriilor dinamice*, în „Revista de statistică" nr. 2/1958.
[99] NEGURĂ, I., *Statistica internațională — disciplină autonomă de cercetare*, în „Revista de statistică" nr. 3/1967.
[100] NEGURĂ, ION, BANU, MIHAELA, SCUTARU-UNGUREANU, CORNELIA, FILIPBEREA, LIANA, *Calcul grafic cu aplicații în agricultură*, Editura „Ceres", București, 1972.

Scanned with OKEN Scanner

[101] NEMCINOV, V. S., *Ekonomico-matematiceskie metodi i modeli*, Izdatelstvo „Mîsl", Moskva, 1965.
[102] NEURATH, P., *Statistik für Sozialwissenschaftler. Eine Einführung in das statistische Denken*, Ferd. Enke Verlag, Stuttgart, 1966.
[103] ONICESCU, O., *Corelația informațională*, în „Revista de statistică" nr. 6/1972.
[104] ONICESCU, O., *Energia informațională — componentă a unui barometru statistic al sistemelor*, în „Revista de statistică" nr. 11, 1968.
[105] ONICESCU, O., *Principiile teoriei probabilităților*, Editura Academiei R.S.R., București, 1969.
[106] ONICESCU, O., *Statistica și sociologia*, în „Arhiva pentru știință și reformă socială" nr. 4/1931.
[107] ONICESCU, O., DEMETRESCU, M. C., *Energia informațională ca măsură a gradului de uniformitate sau de diferențiere a unui sistem*, în „Progresele științei" nr. 2, 1972.
[108] ONICESCU, O., MIHOC, GH., *Lecții de statistică matematică*, Editura tehnică, București, 1956.
[109] PIATIER, A., *Statistique et observation économique*, t. I: Méthodologie statistique, t. II: Econométrie — conjoncture — comptabilité nationale, Presses Universitaires de France, Paris, 1961.
[110] QUÉTELET, A., *Sur l'homme et le développement de ses facultés ou Essai de physique sociales* t. I. Paris, 1835.
[111] RANCU, N., TÖVISSI, L., *Statistica matematică cu aplicații în producție*, Editura Academiei R.S.R., București, 1963.
[112] ROMAN, V., *Cu privire la teoria generală a revoluției științifice și tehnice*, în vol. *Revoluția științifică-tehnică și progresul*, Editura politică, București, 1974.
[113] SĂHLEANU, V., *Metode matematice în cercetarea medico-biologică*, Editura medicală, București, 1957.
[114] SAY, J.-B., *Traité d'économie politique ou simple exposition de la manière dont se forment, se distribuent et se consoment les richesses*, t. I, Paris, 1803.
[115] SCHMIDT, F. C., *Handbook of Graphic Presentation*, The Ronald Press Company, New York, 1954.
[116] SIGMOND, A., *Statistica. Partea generală*, vol. I, Oradea Mare, 1922.
[117] SMIRNOV, I. V., DUNIN-BUKOVSKI, *I. V., Kratki kurs matematiceskoi statistiki dlia tehniceskih prilojenii*, Moskva, 1959.
[118] SUKHATME, D. V., SUKHATME, B. V., *Studii*, în *Sampling Theory of Surveys with Application*, trad. în „Probleme de statistică social-economică" nr. 9—12/1970 și 1—5,7/1971.
[119] SUȚU, N., *Opere economice*, Editura științifică, București, 1957.
[120] SWOBODA, H., *Knaurs Buch der modernen Statistik*, Droemersche Verlagsanstalt Th. Knaur Nachf., München/Zürich, 1971.
[121] ȚICULESCU, D., BIJI E., *Statistica aplicată în economia întreprinderilor agricole socialiste*, Editura „Ceres", București, 1975.
[122] TINTNER, G., *Handbuch der Okonometrie*, Springer-Verlag, Berlin, 1960.
[123] TÖVISSI, L., TREBICI, Vl. *Funcții de producție*, în „Revista de statistică" nr. 12/1972.
[124] TREBICI, Vl., *Mică enciclopedie de demografie*, Editura științifică și enciclopedică, București, 1975.
[125] TREBICI, VL., MOINEAGU, C., *Electrificarea economiei naționale și creșterea productivității muncii sociale*, în „Revista de statistică" nr. 10/1961.
[126] TREBICI, VL., MOINEAGU, C., MARINESCU, I., *Aplicații ale metodei corelației în analiza statistică a productivității muncii*, în „Probleme economice" nr. 4/1962.

Scanned with OKEN Scanner

[127] TUDOR, P., *Istoria sclavajului în Dacia romană*, Editura Academiei R.S.R., București 1957.
[128] TUDOSESCU, I., *Determinismul și știința*, Editura științifică, București, 1971.
[129] URSEANU, V., NEGURĂ, I., *Asupra unor contradicții în teoria indicilor*, în „Studii de statistică", Direcția Centrală de Statistică, 1969.
[130] VENȚELI, E. S., *Teoria veroiatnostei*, Moskva, 1964.
[131] WILKO, S.S., *Mathematical Statistics*, Princeton, 1944, trad. în rusă, Moskva, 1967.
[132] YEOMANS, K. A., *Statistics for the social scientist*, I — Introducing statistics, II — Applied statistics, Penguin Books, Middlesex, 1968.
[133] YULE, G. U., *The applications of the Method of Correlation to Social and Economic Statistics*, în „Bulletin de l'Institut International de Statistique", T. XVIII, Paris, 1909.
[134] YULE, G. U., KENDALL, M. G., *Introducere în teoria statisticii*, Editura științifică, București, 1969.

Scanned with OKEN Scanner

Redactor: MARIA BLAGA
Tehnoredactor: GHEORGHE POPOVICI

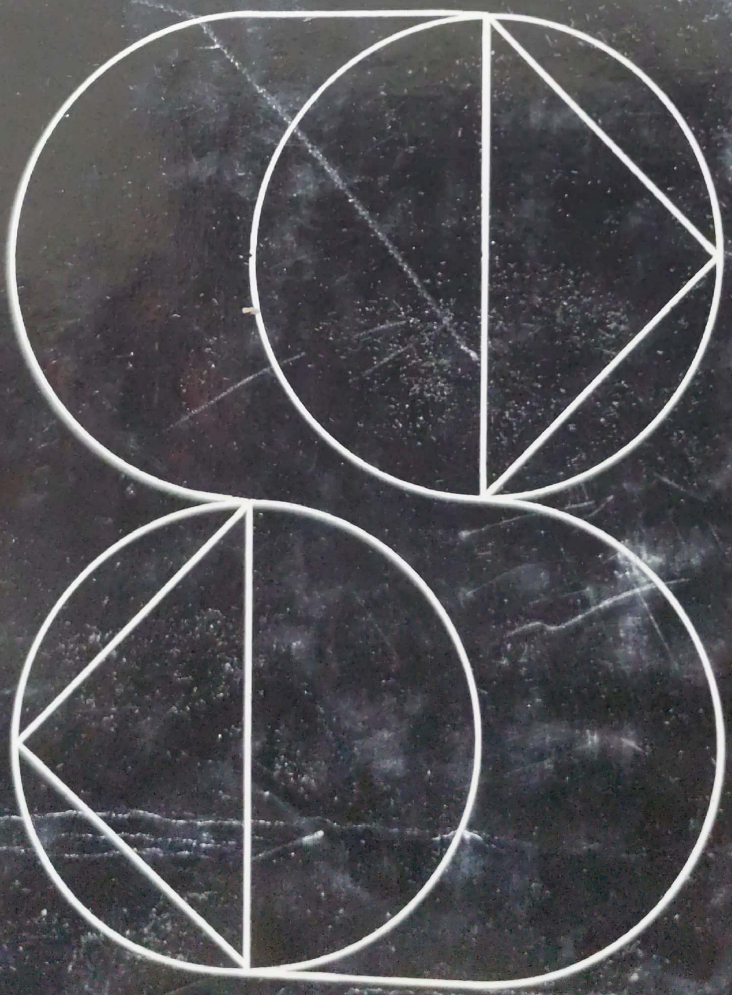
Coli de tipar: 24,25. Tirajul: 4.300 ex. Bun de tipar: 13.XII.1976

Tiparul executat sub comanda
nr. 548 la
Întreprinderea Poligrafică
„13 Decembrie 1918"
str. Grigore Alexandrescu nr. 89—97
București,
Republica Socialistă România

Scanned with OKEN Scanner

Aceea ce mai înainte de toate este necesar statisticianului, în adevăr vrednic de acest nume, este facultatea observației, arta și mai ales tactul de a deosebi aceea ce este adevărat de aceea ce este mincinos în cifrele ce i se înfățișează, probitatea științifică și neînvinsă ce-l face să caute adevărul și numai adevărul...

ION IONESCU DE LA BRAD



Scanned with OKEN Scanner